METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS
PART II: LINEAR STOCHASTIC MODELS

TECHNICAL REPORT NO. 12

T. W. ANDERSON AND N. D. SINGPURWALLA

OCTOBER 1984

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-82-K-1056
THEODORE W. ANDERSON, PROJECT DIRECTOR

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

METHODS AND APPLICATIONS OF TIME SERIES ANALYSIS
PART II: LINEAR STOCHASTIC MODELS

TECHNICAL REPORT NO. 12

T. W. ANDERSON
STANFORD UNIVERSITY

and

N. D. SINGPURWALLA*
THE GEORGE WASHINGTON UNIVERSITY

OCTOBER 1984

U. S. ARMY RESEARCH OFFICE
CONTRACT DAAG29-82-K-0156

Also issued as GWU/IRRA/T-84/11, The Institute for Reliability and Risk Analysis, School of Engineering and Applied Science, George Washington University, Washington D.C. 20052.

*Research Supported by Grant DAAG29-84-K-0160, U.S. Army Research Office, and Contract N00014-77-C-0263, Project NR042-372, Office of Naval Research, with The George Washington University.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA

CONTENTS

Methods and Applications of Time Series Analysis

Part II: Linear Stochastic Models

5. Introduction to Autoregressive Models

In the previous sections we considered models for time series in which the characteristic and useful properties appropriate to the time sequence were embodied in the mean function $f(t)$; $f(t)$ could be a polynomial or a trigonometric function. In astronomy, for example, it is reasonable to suppose that the effect of time is mainly in $f(t)$ and thus prediction is reasonable. In economics and weather, for example, the random part u_t is also time dependent, and thus prediction is more difficult. When the effect of time is embodied in u_t, we are led to a "stochastic process" whose characteristic properties are described by the underlying probabilistic structure. In these cases, for example, there are not regular periodic cycles but more or less irregular fluctuations that have statistical properties of variability. A process whose probability structure does not change with time is called stationary. In Section 5 we are mainly interested in processes that are stationary or almost stationary or such that at least the probability aspect (as distinguished from a deterministic mean value function) is roughly stationary.

To illustrate these ideas, let us consider an autoregressive process of order one, which is described by the relationship

$$y_t = \rho y_{t-1} + u_t \, , \quad t = 1,2,\ldots \, ,$$

where the y_t's are observed values of a random variable, and the u_t's are some unobserved random variables, called innovations. The innovation u_t is assumed independent of $y_{t-1}, y_{t-2}, \ldots$ for all values of t .

The distribution of y_1 and y_2 is given by the distribution of y_1 and $\rho y_1 + u_2$, and similarly the distribution of y_1, y_2, and y_3 is given by the distribution of y_1, $\rho y_1 + u_2$, and $\rho(\rho y_1 + u_2) + u_3$. Thus y_3 depends on y_2, which in turn depends upon y_1, and so on. If $|\rho| < 1$, then the further apart the y's, the less they are related. An innovation u_2 is absorbed into $y_3, y_4, \ldots,$ and thus the randomness perpetuates in time. We therefore say that the effect of time is embodied in the u_t's. The above process is pictorially described in Figure 5.1.

In Section 5.1 we discuss briefly some basic properties of stochastic processes and introduce some notions which are used subsequently.

5.1 Stationary Stochastic Processes

The sequence of T observations which constitute an observed time series may often be considered as a sample at T consecutive equally spaced time points of a much longer sequence of random variables. It is convenient to treat this longer sequence as infinite, extending indefinitely into the future, and possibly going indefinitely into the past. Such a sequence of random variables $y_1, y_2, \ldots,$ or $\ldots, -y_{-2}, -y_{-1}, y_0, y_1, y_2, \ldots,$ is known as a stochastic process with a discrete time parameter. An objective of statistical inference may be to determine the probability structure of the longer infinite sequence.

In a stochastic process those variables that are close together in time generally behave more similarly than those that are far apart

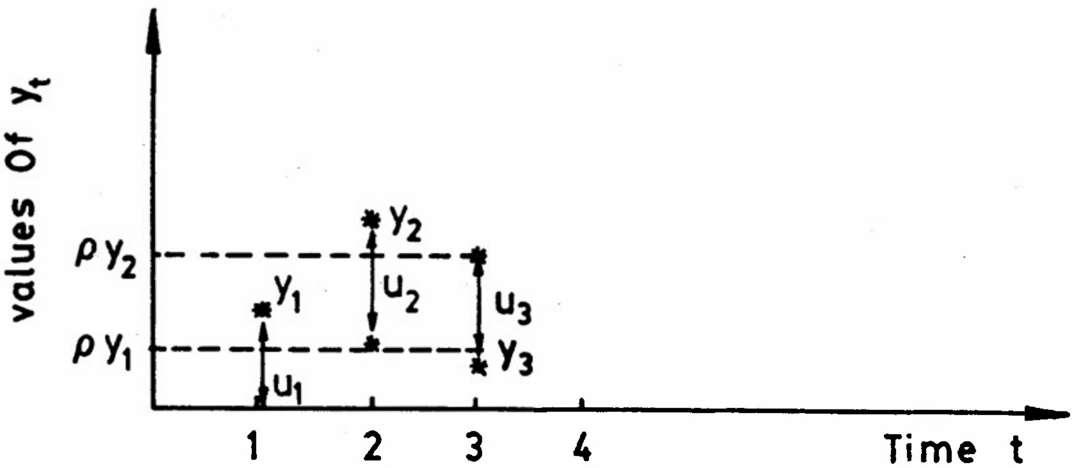

Figure 5.1. An illustration of the structure of an autoregressive process of order 1.

in time. Usually some simplifications are imposed on the probability structure of the larger series, with the result that the finite set of observations has implications for the infinite sequence. One simplifying property is that of stationarity, behind which is the *subjective* idea that the behavior of a set of random variables at one point in time is probabilistically the same as the behavior of a set at another point in time. Thus for example, if the underlying probability structure is assumed to be Gaussian (normal) and stationary, then there is one mean, one variance, and an infinite number of covariances. We are interested in finding out what information about these can be gleaned from a finite number of observations.

A stochastic process $y(t)$ of a continuous time parameter t can be defined for $t \geq 0$ or $-\infty < t < \infty$. A sample from such a process can consist of observations on the process at a finite number of time points, or it can consist of a continuous observation on the process over an interval of time. For example, the sample could be a sequence of consecutive hourly readings of the temperature at some location, or it might be a graph of a continuous reading. Often a stochastic process with a discrete time parameter can be thought of as a sampling at equally spaced time points of a stochastic process of a continuous time parameter.

A discrete time parameter stochastic process is said to be stationary, or strictly stationary, if the distribution of $y_{t_1},\ldots,y_{t_n}$ is the same as the distribution of $y_{t_1+t},\ldots,y_{t_n+t}$ for every finite set of integers $\{t_1,\ldots,t_n\}$ and for every integer t.

We shall denote the mean or the first order moment $\mathcal{E}y_t$ by $m(t)$, and the covariance or the second order moment $\mathcal{E}(y_t - m(t))(y_s - m(s)) = \mathrm{Cov}(y_t, y_s)$ by $\sigma(t,s)$. The sequence $m(t)$ is arbitrary, but the second order moment $\sigma(t,s) = \sigma(s,t)$ for every pair s,t, and the matrix $[\sigma(t_i,t_j)]$, $i,j = 1,\ldots,n$, must be positive semidefinite for every n.

If the first order moments exist, then stationarity implies that

$$\mathcal{E}y_s = \mathcal{E}y_{t+s}\ , \quad s,t = \ldots,-1,0,+1,\ldots\ , \tag{5.1}$$

or that $m(s) = m(s+t) = m$, say, for all s and t. Stationarity also implies that for all $t > 0$ (y_{t_1}, y_{t_2}) has the same distribution as (y_{t_1+t}, y_{t_2+t}), so that if the second moments exist, then

$$\mathrm{Cov}(y_{t_1}, y_{t_2}) = \sigma(t_1,t_2) = \mathrm{Cov}(y_{t_1+t}, y_{t_2+t}) = \sigma(t_1+t\,,\, t_2+t) \quad .$$

If we set $t = -t_2$, then

$$\sigma(t_1,t_2) = \sigma(t_1-t_2,\, 0) = \sigma(t_1-t_2)\ , \quad \text{say} . \tag{5.2}$$

Thus for a stationary process the covariance between any two variables y_t and $y_{t\pm s}$ depends upon s, their distance apart in time. The function $\sigma(s)$ as a function of s, is called the covariance function or the autocovariance function, and the function of s

$$\frac{\mathrm{Cov}(y_t,\, y_{t+s})}{\sqrt{\mathrm{Var}(y_t)}\sqrt{\mathrm{Var}(y_{t+s})}} = \frac{\sigma(s)}{\sqrt{\sigma(0)}\ \sqrt{\sigma(0)}} = \frac{\sigma(s)}{\sigma(0)}\ ,$$

is called the correlation function or the autocorrelation function.

A stochastic process is said to be stationary in the wide sense or weakly stationary if the mean function and the covariance function exist and satisfy (5.1) and (5.2). In the case of the normal distribution, weakly stationary implies stricly stationary and vice versa. In the general case, strictly stationary implies weakly stationary, if the second order moments exist.

5.1.1 Examples of Stationary Stochastic Processes

Example 1: Suppose that the y_t's are independent and identically distributed with

$$\mathcal{E}y_t = m\,, \quad \text{and} \quad \mathrm{Var}(y_t) = \sigma^2\,; \tag{5.3}$$

then

$$\begin{aligned}\sigma(t,s) &= \sigma^2\,, \quad s = t\,,\\ &= 0\,, \quad s \neq t\,.\end{aligned} \tag{5.4}$$

This process is strictly stationary; however, if we drop the requirement of identical distributions, but retain (5.3) and (5.4), then the resulting process is stationary in the wide sense.

Example 2. Suppose that the y_t's are identically equal to a random variable y with

$$\mathcal{E}y_t = m \quad \text{and} \quad \mathrm{Var}\ y_t^2 = \sigma(t,t) = \sigma^2 \ .$$

Then, this process is strictly stationary.

Example 3: Define a sequence of random variables $\{y_t\}$ as follows:

$$(5.5) \qquad y_t = \sum_{j=1}^{q} (A_j \cos \lambda_j t + B_j \sin \lambda_j t) , \quad t = \ldots,-1,0,+1,\ldots ,$$

where the λ_j's are constants such that $0 < \lambda_j < \pi$, and $A_1,\ldots,A_q$, $B_1,\ldots,B_q$ are $2q$ random variables such that

$$\mathcal{E}A_j = \mathcal{E}B_j = 0 , \qquad j=1,\ldots,q ,$$

$$\mathcal{E}A_j^2 = \mathcal{E}B_j^2 = \sigma_j^2 , \qquad j=1,\ldots,q ,$$

$$\mathcal{E}A_iA_j = \mathcal{E}B_iB_j = 0 , \quad i \neq j , \quad i,j=1,\ldots,q, \text{ and}$$

$$\mathcal{E}A_iB_j = 0 , \qquad i,j=1,\ldots,q .$$

Then

$$\mathcal{E}y_t = 0 ,$$

and

$$\begin{aligned}
\mathcal{E}y_ty_s &= \mathcal{E} \sum_{i=1}^{q} \sum_{j=1}^{q} (A_i \cos \lambda_i t + B_i \sin \lambda_i t)(A_j \cos \lambda_j s + B_j \sin \lambda_j s) \\
&= \sum_{j=1}^{q} [\mathcal{E}A_j^2 \cos \lambda_j t \cos \lambda_j s + \mathcal{E}B_j^2 \sin \lambda_j t \sin \lambda_j s] \\
&= \sum_{j=1}^{q} \sigma_j^2 [\cos \lambda_j t \cos \lambda_j s + \sin \lambda_j t \sin \lambda_j s] \\
&= \sum_{j=1}^{q} \sigma_j^2 \cos \lambda_j (t-s) .
\end{aligned}$$

Since the covariance of (y_t, y_s) depends only on $(t-s)$, the distance between the two observations, and since $\mathcal{E}y_t = 0$ for all t, the sequence $\{y_t\}$ is stationary in the wide sense. If, however, the A_j's and the B_j's are also normally distributed, then the y_t's will also be normally distributed, and then the process will be stationary in the strict sense.

The point of this example is that every weakly stationary process can be approximated by a linear combination of the type indicated by (5.5).

Example 4: Let $\ldots, v_{-1}, v_0, v_1, \ldots$ be a sequence of independent and identically distributed random variables, and let $\alpha_0, \alpha_1, \ldots, \alpha_q$, be $q+1$ coefficients. Then

(5.6) $$y_t = \alpha_0 v_t + \alpha_1 v_{t-1} + \ldots + \alpha_q v_{t-q}, \quad t = \ldots, -1, 0, 1, \ldots ,$$

is a stationary stochastic process. If $\mathcal{E}v_t = \gamma$, and $\operatorname{Var} v_t = \sigma^2$, then

$$\mathcal{E}y_t = \gamma(\alpha_0 + \alpha_1 + \ldots + \alpha_q)$$

and

$$\begin{aligned} \operatorname{Cov}(y_t, y_{t+s}) &= \sigma^2(\alpha_0\alpha_s + \ldots \alpha_{q-s}\alpha_q), && s = 0, \ldots, q, \\ &= 0, && s = q+1, \ldots, \end{aligned}$$

and so $\{y_t\}$ is weakly stationary. Thus, for $\{y_t\}$ to be weakly stationary, all we need is that the v_t's have the same mean, the same variance, and that they be uncorrelated.

The process (5.6) is known as a finite moving average.

<u>The infinite moving average</u>

$$y_t = \sum_{s=0}^{\infty} \alpha_s v_{t-s} \tag{5.7}$$

means that the random variable y_t, when it exists, is such that

$$\lim_{n\to\infty} \mathcal{E}(y_t - \sum_{s=0}^{n} \alpha_s v_{t-s})^2 = 0 \; . \tag{5.8}$$

A sufficient condition for the existence of y_t is that the v_t's be uncorrelated with a common mean $(=0)$ and variance, and

$$\sum_{s=0}^{\infty} \alpha_s^2 < \infty \; ; \tag{5.9}$$

see Anderson (1971), p.377.

When (5.8) holds, the infinite sum $\sum_{s=0}^{\infty} \alpha_s v_{t-s}$ is said to <u>converge in the mean</u> or in the <u>quadratic mean</u>.

6. <u>Basic Notions of Multivariate Normal Distributions</u>

Two random variables X and Y with means μ_x and μ_y and variances σ_x^2 and σ_y^2, respectively, are said to have a <u>bivariate normal distribution</u>, or a <u>bivariate Gaussian distribution</u>, if their joint density function is given by

$$f(x,y) = \frac{1}{\sigma_x\sigma_y 2\pi\sqrt{1-\rho_{xy}^2}} \exp\left\{-\frac{1}{2(1-\rho_{xy}^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2+\left(\frac{y-\mu_y}{\sigma_y}\right)^2-2\rho_{xy}\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right)\right]\right\} ;$$

then $\rho_{xy} = \dfrac{\mathcal{E}(X-\mu_x)(Y-\mu_y)}{\sigma_x\sigma_y}$ is the <u>correlation</u> between X and Y ; $-\infty < x < \infty$, $-\infty < y < \infty$.

The <u>marginal density</u> of X is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\;\sigma_x} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu_x}{\sigma_x}\right)^2\right\} , \qquad -\infty < x < \infty .$$

This is the normal density function, which we will henceforth denote by $n(x|\mu_x, \sigma_x^2)$.

Similarly, the density function of Y is also normal, $n(y|\mu_y, \sigma_y^2)$.

We can show (Anderson (1984), p.37), that $f(x|y^*)$, the conditional density function of X , given $Y = y^*$ is also normal, but with a mean $\mu_x + \rho_{xy} \frac{\sigma_x}{\sigma_y} (y^*-\mu_y)$, and variance $\sigma_x^2(1-\rho_{xy}^2)$. That is,

$$f(x|y^*) = n(x|\mu_x+\rho_{xy} \frac{\sigma_x}{\sigma_y} (y^*-\mu_y) , \quad \sigma_x^2(1-\rho_{xy}^2)) .$$

Thus, the variance of X given that $Y = y^*$ does not depend on y^* , and its mean is a linear function of y^* .

The mean value of a variate in a conditional distribution, when regarded as a function of the fixed variate, is called a <u>regression</u>. Thus, the regression of X in the situation above is $\mu_x + \rho_{xy} \frac{\sigma_x}{\sigma_y} (y^*-\mu_y)$.

The <u>trivariate normal distribution</u> of three random variables X,Y, and Z is defined in a manner akin to the bivariate normal distribution,

once the means μ_x , μ_y , and μ_z , the variances σ_x^2, σ_y^2, and σ_z^2, respectively, and the correlations between their pairs ρ_{xy}, ρ_{xz}, and ρ_{yz} , are specified. Let $f(x,y,z)$ denote the joint density function of this trivariate normal distribution; let $f(x,y|z)$ denote the joint density function of X and Y conditional on $Z = z$. Then

$$f(x,y|z) = \frac{f(x,y,z)}{f(z)} ,$$

where $f(z) > 0$ is the marginal density function of Z , which is again normal. A property of the normal distributions is that $f(x,y|z)$ is a bivariate normal density (Anderson (1984), p.37).

Let $f(x|z)$ and $f(y|z)$ denote the marginal densities of X and Y conditional on z , respectively; these densities can be obtained via $f(x,y|z)$. Let $\mathcal{E}(X|z)$ and $\mathcal{E}(Y|z)$ denote the expected values of X and Y conditional on z , respectively. From our previous discussions on the bivariate case, we recall that $f(x|z)$ and $f(y|z)$ are also normal, and that $\mathcal{E}(X|z)$ and $\mathcal{E}(Y|z)$ can be written as

$$\mathcal{E}(X|z) = \alpha + \beta z , \quad \text{and}$$
$$\mathcal{E}(Y|z) = \gamma + \delta z .$$

The correlation between X and Y conditional on z , denoted by $\rho_{xy \cdot z}$ is called the <u>partial correlation</u> between X and Y when Z is held constant.

Thus, we have

$$\rho_{xy \cdot z} = \frac{\mathcal{E}(X-(\alpha+\beta z))\mathcal{E}(Y-(\gamma+\delta z))}{\sqrt{\mathcal{E}(X-(\alpha+\beta z))^2 \mathcal{E}(Y-(\gamma+\delta z))^2}} .$$

Small values of $\rho_{xy \cdot z}$ imply that there is little relationship between X and Y that is not explained by Z. We can also verify [Anderson (1984), p.41) that

$$\rho_{xy \cdot z} = \frac{\rho_{xy} - \rho_{xz}\ \rho_{yz}}{\sqrt{1-\rho_{xz}^2}\ \sqrt{1-\rho_{yz}^2}} .$$

To discuss the idea of the "multiple correlation" between X and the pair (Y,Z), let us denote by $\mathcal{E}(X|y,z)$ the expected value of X conditional upon $Y=y$, and $Z=z$. Again, from our discussion of the bivariate case, we note that $\mathcal{E}(X|y,z)$ can be written as

$$\mathcal{E}(X|y,z) = \alpha + \theta y + \gamma z$$

where α, θ, and γ are constants.

Now let us consider the correlation between X and an arbitrary linear combination of Y and Z, say $bY + cZ$, where b and c are arbitrary constants.

Then, the <u>multiple correlation</u> between X and (Y,Z), say R^2 is

$$R^2 = \max_{b,c}[\text{Correlation}(X, (bY+cZ))]^2 .$$

It turns out that the values of b and c are θ and γ respectively. Thus, the multiple correlation is the correlation between X and $\theta Y + \gamma Z$.

7. Estimation of the Correlation Function

One of the first steps in analyzing a time series is to decide whether the observations $y_1, y_2, \ldots, y_T$ are from a process of

independent random variables or from one in which the successive variables are correlated. If the process is assumed stationary, then r(h), an estimate of the correlation function, enables us to infer the nature of the joint distribution that generates the T observations. To see this, consider a pair of random variables Y_t and Y_{t+k}, separated by some lag k, where $k=1,2,\ldots$. The nature of their joint probability distribution can be inferred by plotting a "scatter diagram" using the pair of values y_t and y_{t+k}, for $t=1,2,\ldots,T-k$.

In Figure 7.1 we show a scatter diagram for Y_t and Y_{t+k}; this diagram indicates that a large value of Y_t tends to lead us to a large value of Y_{t+k}, and vice versa. When this happens, we say that Y_t and Y_{t+k} are positively correlated. In Figure 7.2 the scatter diagram shows that a large value of Y_t leads us to a small value of Y_{t+k} and vice versa; in this case, we say that Y_t and Y_{t+k} are negatively correlated. A key requirement underlying our ability to plot and interpret the scatter diagram is the assumption of stationarity. Because of this assumption the joint distribution of Y_t and Y_{t+k} is the same as the joint distribution of any other pair of random variables separated by a lag of k, say Y_{t+s} and Y_{t+s+k}, for some $s \neq 0$.

A formal way of describing the impressions conveyed by a scatter plot is via an estimate of the correlation function; this estimate is also known as the serial correlation. If the observations $y_1, y_2, \ldots, y_T$ are assumed to be generated by a process with mean 0, then $r^*(1)$, the first order serial correlation coefficient is defined as

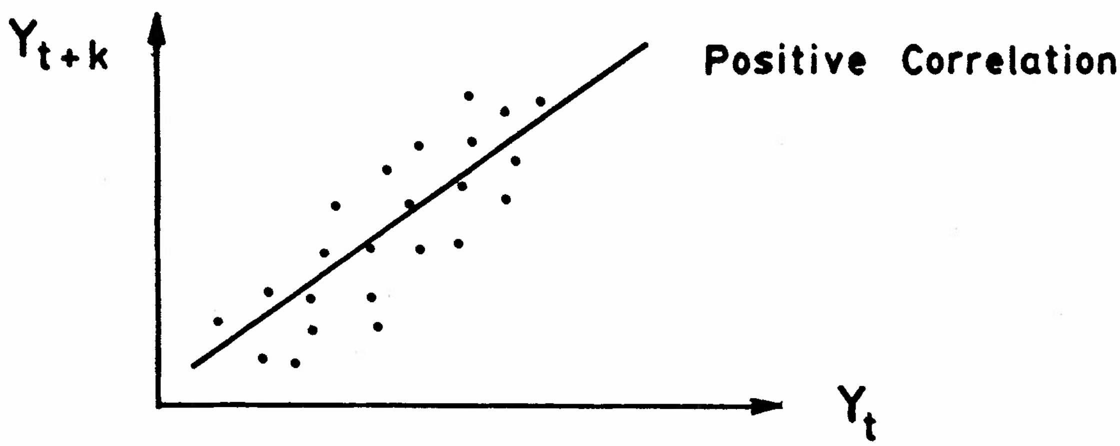

Figure 7.1. Scatter plot of Y_t and Y_{t+k} showing a positive correlation between the variables.

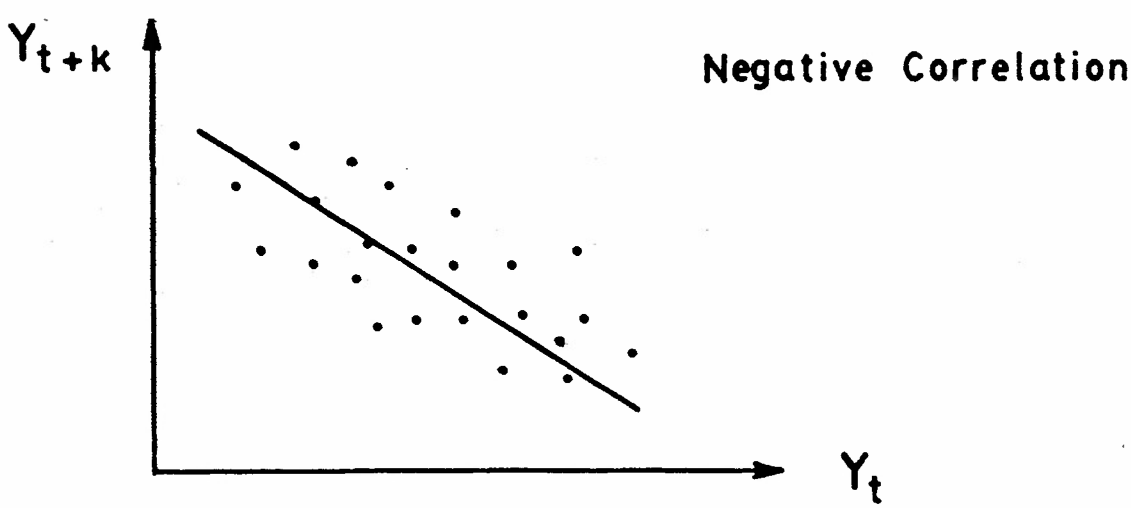

Figure 7.2. Scatter plot of Y_t and Y_{t+k} showing a negative correlation between the variables.

$$(7.1) \qquad r^*(1) = \frac{\sum_{t=1}^{T-1} y_t y_{t+1}}{\sum_{t=1}^{T} y_t^2} .$$

If the mean of the process is not known, (7.1) is modified by replacing y_t and y_{t-1} by the deviation of these from the sample mean $\bar{y}$, where $\bar{y} = \sum_{t=1}^{T} y_t/T$. Thus we have

$$(7.2) \qquad r(1) = \sum_{t=1}^{T-1} (y_t-\bar{y})(y_{t+1}-\bar{y}) / \sum_{t=1}^{T} (y_t-\bar{y})^2 .$$

Higher order serial correlations are similarly defined; for example, $r^*(h)$, the <u>h-th order serial correlation</u> is

$$(7.3) \qquad r^*(h) = \frac{\sum_{t=1}^{T-h} y_t \, y_{t+h}}{\sum_{t=1}^{T} y_t^2} ,$$

or in analogy with (7.2) it is

$$(7.4) \qquad r(h) = \frac{\sum_{t=1}^{T-h} (y_t-\bar{y})(y_{t+h}-\bar{y})}{\sum_{t=1}^{T} (y_t-\bar{y})^2} .$$

8. Autoregressive Processes

One of the simplest, and perhaps the most useful, stochastic process which is used to model a time series is the autoregressive process. A sequence of random variables $y_1, y_2, \ldots$ is said to be an autoregressive process of order p, abbreviated as AR(p), if for some constant μ and integer p

$$(8.1) \qquad (y_t-\mu)+\beta_1(y_{t-1}-\mu)+\ldots+\beta_p(y_{t-p}-\mu)=u_t\,, \quad t=p+1,p+2,\ldots\,,$$

with $u_{p+1}, u_{p+2},\ldots,$ being independent and identically distributed with mean 0 and variance σ^2, and u_t independent of $y_{t-1}, y_{t-2}, \ldots$. We shall set $\mu=0$ in the following discussion. The random variable u_t is called an innovation or a disturbance. We shall refer to the sequence $\{u_t\}$ as an innovation process.

It is convenient to generalize (8.1) to a doubly infinite sequence $\ldots, y_{-1}, y_0, y_1, \ldots,$ resulting in a doubly infinite sequence $\ldots, u_{-1}, u_0, u_1, \ldots$. Such processes are also known as autoregressive processes.

If we use the foreward lag operator $\wp$, where $\wp^k u_t \overset{\text{def}}{=} u_{t+k}$ for any integer k, then (8.1) can also be written as

$$(8.2) \qquad (\wp^p+\beta_1\wp^{p-1}+\ldots+\beta_p\wp^0)y_{t-p}=u_t\,.$$

Since $\Delta y_t = y_{t+1} - y_t = \wp y_t - y_t = (\wp-1)y_t$, we have the result that $\Delta = \wp-1$; recall that Δ is the foreward difference operator introduced in Section 3.3. Thus we may say that the operator acting on y_{t-p} can also be written as a polynomial in Δ of degree p. If $\beta_p \neq 0$, then the left hand side of (8.2) can be written as a linear combination of $y_{t-p}, \Delta y_{t-p}, \Delta^2 y_{t-p}, \ldots, \Delta^p y_{t-p}$ and is therefore called a stochastic difference equation of degree p.

Unless otherwise stated (see for instance Section 8.9), we shall assume that the stochastic process described by (8.2) is stationary. In Section 8.1 we shall determine the conditions under which u_t is independent of $y_{t-1}, y_{t-2}, \ldots,$.

The model (8.1) can be used to generate other processes. For example, should we want to incorporate the effect of a trend in (8.1), then we add to the left hand side of (8.1) the term $\sum_i \gamma_i z_{it}$, where the z_{it}'s are known functions of time; this matter is discussed further in Section 8.7.

Autoregressive processes were suggested by Yule (1927), and were applied by him to study sunspot data. Gilbert Walker (1931) extended the theory and applied it to atmostpheric data. In what follows we shall study the structure of autoregressive processes, and address the related questions of inference and prediction.

8.1 Representation as an Infinite Moving Average

If we inspect (8.1), we see that y_t is expressed as a linear combination of the previous y_t's and u_t. We shall now study the conditions under which y_t can be written as an infinite linear combination of u_t and the earlier u_r's. To see the idea, we consider an AR(1) process

$$y_t = \rho y_{t-1} + u_t ,$$

and note that since $y_{t-1} = \rho y_{t-2} + u_{t-1}$, we have

$$y_t = u_t + \rho u_{t-1} + \rho^2 y_{t-2} .$$

Successive substitution of the type indicated above leads us to write

(8.3) $$y_t = u_t + \rho u_{t-1} + \rho^2 u_{t-2} + \ldots + \rho^s u_{t-s} + \rho^{s+1} y_{t-(s+1)}$$

so that

(8.4) $$y_t - (u_t + \rho u_{t-1} + \ldots + \rho^s u_{t-s}) = \rho^{s+1} y_{t-(s+1)} .$$

The difference between y_t and a linear combination of the $(s+1)$ u_r's is therefore $\rho^{s+1} y_{t-(s+1)}$, and this becomes small when $|\rho| < 1$ and s is large. In particular

(8.5) $$\mathcal{E}[y_t - (u_t + \rho u_{t-1} + \ldots + \rho^s u_{t-s})]^2 = \rho^{2(s+1)} \mathcal{E} y^2_{t-(s+1)}$$

will not depend on t, if we assume that $\{y_t\}$ is a doubly infinite stationary process. As s increases, (8.5) will go to 0, and so we can write

$$y_t = \sum_{r=0}^{\infty} \rho^r u_{t-r}$$

and say that the infinite sum on the right of the above equation <u>converges in the mean</u> to y_t. (See Section 5.7.)

Let us now consider an AR(p),

$$\sum_{r=0}^{p} \beta_r y_{t-r} = u_t , \qquad \beta_0 = 1 .$$

so that

$$y_t = u_t - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \ldots - \beta_p y_{t-p} .$$

Replacement of t by t-1 yields

$$y_{t-1} = u_{t-1} - \beta_1 y_{t-2} - \beta_2 y_{t-3} - \dots - \beta_p y_{t-p-1} ,$$

which upon substitution gives

$$y_t = u_t - \beta_1(u_{t-1} - \beta_1 y_{t-2} - \beta_2 y_{t-3} - \dots - \beta_p y_{t-p-1}) - \beta_2 y_{t-2} - \dots - \beta_p y_{t-p}$$

$$= u_t - \beta_1 u_{t-1} - (\beta_2 - \beta_1^2) y_{t-2} - \dots + \beta_1 \beta_p y_{t-1-p} .$$

Continuing in the above manner s times, we arrive at

$$y_t = u_t + \delta_1^* u_{t-1} + \dots + \delta_s^* u_{t-s} + \alpha_{s1}^* y_{t-s-1} + \alpha_{s2}^* y_{t-s-2} + \dots + \alpha_{sp}^* y_{t-s-p} . \tag{8.6}$$

We note that each substitution leaves us with p consecutive y_r's on the right-hand side of the above. Since $y_{t-s-1} = u_{t-s-1} - \beta_1 y_{t-s-2} - \dots - \beta_p y_{t-s-p-1}$, we have

$$y_t = u_t + \delta_1^* u_{t-1} + \dots + \delta_s^* u_{t-s} + \alpha_{s1}^*(u_{t-s-1} - \beta_1 y_{t-s-2} \dots - \beta_p y_{t-s-p-1})$$

$$+ \alpha_{s2}^* y_{t-s-2} + \dots + \alpha_{sp}^* y_{t-s-p} .$$

$$= u_t + \delta_1^* u_{t-1} + \dots + \delta_s^* u_{t-s} + \alpha_{s1}^* u_{t-s-1} + (\alpha_{s2}^* - \alpha_{s1}^* \beta_1) y_{t-s-2}$$

$$+ \dots + (\alpha_{sp}^* - \alpha_{s1}^* \beta_{p-1}) y_{t-s-p} - \alpha_{s1}^* \beta_p y_{t-s-p-1} .$$

Thus

$$\delta_{s+1}^* = \alpha_{s1}^* ,$$

$$\alpha_{s+1,j}^* = (\alpha_{s,j+1}^* - \alpha_{s1}^* \beta_j) , \qquad j = 1, \dots, p-1 ,$$

$$\alpha_{s+1,p}^* = -\alpha_{s1}^* \beta_p$$

is a set of recursion relationships for the coefficients. Continuation of this procedure leads us to write, for $\delta_0^* = 1$,

$$(8.7) \qquad y_t = \sum_{i=0}^{\infty} \delta_i^* u_{t-i}$$

<u>if</u> the infinite sum on the right-hand side of (8.7) converges in the mean to y_t. We shall next see the conditions for this convergence.

8.1.1 <u>Conditions for Convergence in the Mean of Autoregressive Processes</u>

The material of Section 8.1 can be formalized by using the <u>backward lag operator</u> $\mathcal{L}$, where $\mathcal{L}y_t = y_{t-1}$, and writing the process (8.1) as

$$\sum_{r=0}^{p} \beta_r \mathcal{L}^r y_t = u_t \, .$$

Then, formally we can write our AR(p) process as

$$y_t = \left(\sum_{r=0}^{p} \beta_r \mathcal{L}^r\right)^{-1} u_t \, ,$$

where

$$\left(\sum_{r=0}^{p} \beta_r \mathcal{L}^r\right)^{-1} = \sum_{r=0}^{\infty} \delta_r \mathcal{L}^r \, ;$$

the δ_r's are the coefficients in the equality

$$(8.8) \qquad (\sum_{r=0}^{p} \beta_r z^r)^{-1} = \sum_{r=0}^{\infty} \delta_r z^r$$

on the basis that the above equality can be so written meaningfully.

It can be verified (Anderson (1971), p.169) that the δ_r's of (8.8) are indeed the same as the δ_r^*'s of (8.7), which we recall were obtained by successive substitution; thus we write $\delta_r = \delta_r^*$.

In order to see the conditions under which it is meaningful to write (8.8), we consider

$$(8.9) \qquad \beta_0 x^p + \beta_1 x^{p-1} + \ldots + \beta_p x^0 = 0 ,$$

the <u>associated polynomial equation</u> of the stochastic difference equation (8.1) (our AR(p) process).

For $\beta_p \neq 0$, let $x_1,\ldots,x_p$ be the p roots of (8.9). If $|x_i| < 1$, for $i=1,\ldots,p$, then it is clear that $z_1,\ldots,z_p$, the roots of

$$\sum_{r=0}^{p} \beta_r z^r = 0 ,$$

are such that $z_i = 1/x_i$ and that $|z_i| > 1$. Now, for any z such that $|z| < \min_i |z_i|$, the series

$$(8.10) \qquad \frac{1}{\sum_{r=0}^{p} \beta_r z^r} = \frac{1}{\prod_{i=1}^{p} (1 - \frac{z}{z_i})} = \prod_{i=1}^{p} \sum_{r=0}^{\infty} (\frac{z}{z_i})^r = \sum_{r=0}^{\infty} \delta_r z^r ,$$

converges absolutely. Thus we see that when $x_1,\ldots,x_p$, the roots of the associated polynomial equation of an AR(p) process, are less than 1 in absolute value, we can write $\left(\sum_{r=0}^{p} \beta_r z^r\right)^{-1} = \sum_{r=0}^{\infty} \delta_r z^r$.

To argue convergence in the mean of the AR(p) process, we consider the expression $\left(\sum_{r=0}^{p} \beta_r z^r\right)^{-1}$ and note that by a formal long hand division

$$\frac{1}{1+\beta_1 z+\ldots+\beta_p z^p} = 1 - \frac{\beta_1 z+\beta_2 z^2+\ldots+\beta_p z^p}{1+\beta_1 z+\beta_2 z^2+\ldots+\beta_p z^p}$$

$$= 1 - \beta_1 z - \frac{(\beta_2-\beta_1^2)z^2+\ldots+(\beta_p-\beta_1\beta_{p-1})z^p-\beta_1\beta_p z^{p+1}}{1+\beta_1 z+\beta_2 z^2+\ldots+\beta_p z^p} .$$

If we continue in the above manner, we see that

$$\frac{1}{1+\beta_1 z+\ldots+\beta_p z^p} = 1+\delta_1 z+\delta_2 z^2+\ldots+\delta_s z^s + \frac{\alpha_{s1} z^{s+1}+\ldots+\alpha_{sp} z^{s+p}}{1+\beta_1 z+\ldots+\beta_p z^p} ,$$

where the δ_r's and the α_{si}'s satisfy the same recurrence relationships as the δ_r^*'s and the α_{si}^*'s of Section 8.1. Thus $\delta_r^* = \delta_r$ and $\alpha_{si}^* = \alpha_{si}$. (See 8.6.)

In view of (8.10), we now see that $\dfrac{\alpha_{s1} z^{s+1}+\ldots+\alpha_{sp} z^{s+p}}{1+\beta_1 z+\ldots+\beta_p z^p}$ must converge to 0 for $|z| < \min_i |z_i|$, and in particular for $z = 1$. This implies that the $\alpha_{si} \to 0$ (as $s \to \infty$) for each i . Thus, if $\{y_t\}$ is

a stationary process

$$\mathcal{E}(y_t - \sum_{r=0}^{s} \delta_r u_{t-r})^2 = \mathcal{E}(\alpha_{s1} y_{t-s-1} + \cdots + \alpha_{sp} y_{t-s-p})^2$$

will not depend on t and will converge to 0 as $s \to \infty$.

We therefore have

$$y_t = \sum_{r=0}^{\infty} \delta_r u_{t-r}$$

in the sense of convergence in the mean. We have proved

Theorem 8.1: If the roots of the polynomial equation $\sum_{r=0}^{p} \beta_r x^{p-r} = 0$ associated with a stationary AR(p) process $\sum_{r=0}^{p} \beta_r y_{t-r} = u_t$ are less than 1 in absolute value, then y_t can be written as an infinite linear combination of $u_t, u_{t-1}, u_{t-2}, \ldots,$.

Note that whenever y_t can be written as an infinite linear combination of $u_t, u_{t-1}, \ldots,$ y_t will be independent of the future innovations $u_{t+1}, u_{t+2}, \ldots,$; this follows from our assumption that the sequence of innovations $\{u_t\}$ is mutually independent. We thus have as a corollary to Theorem 8.1

Corollary 8.2: If the roots of the polynomial equation associated with a stationary AR(p) process are less than 1 in absolute value, y_t is independent of $u_{t+1}, u_{t+2}, \ldots,$.

8.2 Evaluation of the Coefficients δ_r and their Behavior

Suppose that the roots of the associated polynomial equation $\sum_{r=0}^{p} \beta_r x^{p-r} = 0$, are less than 1 in absolute value. Then, by Theorem 8.1, we can write

$$y_t = \sum_{r=0}^{\infty} \delta_r u_{t-r} ,$$

where the δ_r's are to be viewed as weights associated with the present and past innovations $u_t, u_{t-1}, \dots$. Our goal is to determine a procedure by which the δ_r's can be expressed in terms of the known β_r's, and also to see if there is any discernable pattern in the δ_r's. Such a pattern will enable us to interpret the behavior of our sequence $\{y_t\}$.

From (8.10) we note that since

$$\frac{1}{\sum_{r=0}^{p} \beta_r z^r} = \sum_{r=0}^{\infty} \delta_r z^r ,$$

$$\left(\sum_{r=0}^{p} \beta_r z^r\right)^{-1} \sum_{s=0}^{p} \beta_s z^s = \left(\sum_{r=0}^{\infty} \delta_r z^r\right) \sum_{s=0}^{p} \beta_s z^s = 1 ,$$

or that

$$\sum_{r=0}^{\infty} \sum_{s=0}^{p} \beta_s \delta_r z^{r+s} = 1 .$$

Replacing r by $(t-s)$ and by suitable re-arrangements, we have

$$\sum_{t=0}^{p-1} \left(\sum_{s=0}^{t} \beta_s \delta_{t-s} \right) z^t + \sum_{t=p}^{\infty} \left(\sum_{s=0}^{p} \beta_s \delta_{t-s} \right) z^t = 1 ,$$

which is an identity in z (the series converging absolutely for $|z| \le 1$). An inspection of the above reveals that the coefficient of

z^0 on the left hand side is 1 and the coefficients of the other powers of z are zero; thus we have the following set of relationships between the δ's and the β's :

$$(8.11)\qquad \begin{cases} \beta_0\delta_0 = \delta_0 = 1 , \\ \beta_0\delta_1 + \beta_1\delta_0 = \delta_1 + \beta_1 = 0 , \\ \vdots \\ \beta_0\delta_{p-1} + \ldots + \beta_{p-1}\delta_0 = 0 , \end{cases} \qquad \text{and}$$

$$(8.12)\qquad \beta_0\delta_t + \ldots + \beta_p\delta_{t-p} = 0 , \qquad t = p,\ p+1, \ldots, .$$

We note that (8.12) is a homogeneous difference equation which corresponds to the (stochastic) difference equation that describes the AR(p) process

$$\beta_0 y_t + \beta_1 y_{t-1} + \ldots + \beta_p y_{t-p} = u_t , \qquad \beta_0 = 1 .$$

If the roots of (8.9), the polynomial equation associated with an AR(p) process are distinct, then the general solution of (8.12) is of the form

$$(8.13)\qquad \delta_r = \sum_{i=1}^{p} k_i x_i^r \qquad r = 0,1,\ldots,$$

where $k_1,\ldots,k_p$ are coefficients.

If a root x_i is real, then the coefficient k_i is also real. If a pair of roots x_i and x_{i+1} are conjugate complex, then k_i and k_{i+1} are also conjugate complex and $k_i x_i^r + k_{i+1} x_{i+1}^r$ is real, $r = 0,1,\ldots$.

Equations (8.11) give us the boundary conditions for solving (8.12). The p equations (8.11) enable us to determine the p constants $k_1,\ldots,k_p$ by substituting (8.13) in (8.11).

The above material can be better appreciated via some special cases; these are discussed below.

8.2.1 Special Cases Describing the Evaluation and Behavior of δ_r's

We shall consider here two examples, an autoregressive process of order 1 and an autoregressive process of order 2.

An Autoregressive Process of Order 1

Suppose that in (8.1) $p = 1$ (and $\mu = 0$), so that an AR(1) process is

$$\beta_0 y_t + \beta_1 y_{t-1} = u_t , \qquad \text{for} \quad t = 2,3,\ldots \;.$$

The associated polynomial equation for the above process is

$$\beta_0 x + \beta_1 x^0 = 0 ,$$

and so with $\beta_0 = 1$, $x = -\beta_1$ is the only root.

From (8.11) we have $\delta_0 = 1$ and $\delta_1 = -\beta_1$, so that the coefficient k_1 in (8.13) is 1. Thus, for our AR(1) process the coefficients δ_r are such that

$$\delta_r = k_1 x_1^r = (-\beta_1)^r \quad . \tag{8.14}$$

Now, if we assume that the process is stationary, then in order to be able to write y_t as an infinite linear combination of $u_t, u_{t-1}, \ldots,$ we need to have, by Theorem 8.1, $|x| < 1$ or equivalently $|\beta_1| < 1$. Thus, when $|\beta_1| < 1$, we can write

$$y_t = \sum_{r=0}^{\infty} \delta_r u_{t-r} \; . \tag{8.15}$$

We note from (8.14), that the weights δ_r exponentially decay in r when $|\beta_1| < 1$. The decay is smooth if $\beta_1 < 0$, and the decay alternates in sign if $\beta_1 > 0$. This behavior of the δ_r's implies that in (8.15) the remote innovations receive smaller weights than the more recent ones. Such results are useful for explaining the behavior of the series $y_t, y_{t-1}, \ldots,$ and also interpreting forecasts in autoregressive processes or order 1.

An Autoregressive Process of Order 2

Now suppose that in (8.2) $p = 2$ (and $\mu = 0$), so that an AR(2) process is

$$\beta_0 y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} = u_t \; , \qquad t = 3,4,\ldots \quad .$$

With $\beta_0 = 1$ the associated polynomial equation for our AR(2) process becomes

$$x^2 + \beta_1 x + \beta_2 x^0 = 0 .$$

If x_1 and x_2 are the roots of the above equation, then $x_i = (-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_2})/2$, $i = 1,2$.

If the roots x_1 and x_2 are real and distinct, that is $\beta_1^2 > 4\beta_2$, then (8.11) and (8.12) give $1 = k_1 x_1^0 + k_2 x_2^0 = k_1 + k_2$ and $k_1 x_1 + k_2 x_2 = -\beta_1 = x_1 + x_2$. The solution is

$$k_1 = \frac{x_1}{x_1 - x_2} \quad \text{and} \quad k_2 = \frac{-x_2}{x_1 - x_2} .$$

Then

$$(8.16) \qquad \delta_r = \frac{x_1^{r+1} - x_2^{r+1}}{x_1 - x_2}, \qquad r = 0,1,2,\ldots .$$

If we assume that our AR(2) process is stationary, then in order to be able to write y_t in the form (8.15), that is, as an infinite linear combination of $u_t, u_{t-1}, \ldots,$ we need to have (by Theorem 8.1) $|x_i| < 1$, $i = 1,2$. This in turn implies that the coefficients β_1 and β_2 will have to satisfy the following conditions:

$$(8.17) \qquad \begin{aligned} &\beta_1 + \beta_2 > -1, \\ &\beta_1 - \beta_2 < 1, \quad \text{and} \\ &-1 < \beta_2 < 1 . \end{aligned}$$

The above conditions define a triangualr region, shown in Figure 8.1, in which the coefficients β_1 and β_2 must lie; also see Box and Jenkins, (1976), p.59.

When $|x_i| < 1$, $i = 1,2$, and x_1 and x_2 are real, that is, when β_1 and β_2 lie outside the parabolic region of Figure 8.1, then from (8.16) it is clear that the weights δ_r are a linear combination of two exponentially decaying functions of r, x_1^{r+1} and x_2^{r+1}.

When $|x_i| < 1$, $i = 1,2$, and when x_1 and x_2 are complex, that is $\beta_1^2 < 4\beta_2$ so that β_1 and β_2 lie in the parabolic region of Figure 8.1, x_1 and x_2 may be written as $x_1 = \alpha e^{i\theta}$ and $x_2 = \alpha e^{-i\theta}$, where $i = \sqrt{-1}$; since $|x_1| < 1$ and $|x_2| < 1$, $\alpha < 1$. Thus

$$k_1 = \frac{e^{i\theta}}{e^{i\theta} - e^{-i\theta}} \quad \text{and} \quad k_2 = \frac{-e^{-i\theta}}{e^{i\theta} - e^{-i\theta}} ,$$

so that

$$\delta_r = k_1 x_1^r + k_2 x_2^r = \alpha^r \frac{e^{i\theta(r+1)} - e^{-i\theta(r+1)}}{e^{i\theta} - e^{-i\theta}} = \alpha^r \frac{\sin(\theta(r+1))}{\sin\theta} , \tag{8.18}$$

since $e^{i\theta} = \cos\theta + i \sin\theta$.

Thus δ_r is a damped sine function of r, whose nature is illustrated in Figure 8.2. Such a damped sinusoidal behavior of the weights offers an explanation of an oscillatory pattern of the y_t's often observed in otherwise nonperiodic stationary time series. (Also see Section 8.6.)

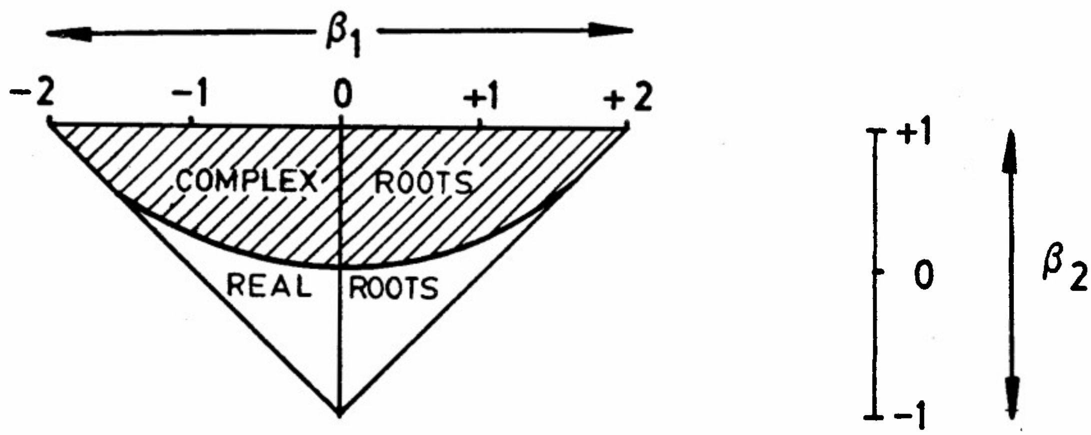

Figure 8.1. Region defining admissible values of β_1 and β_2 .

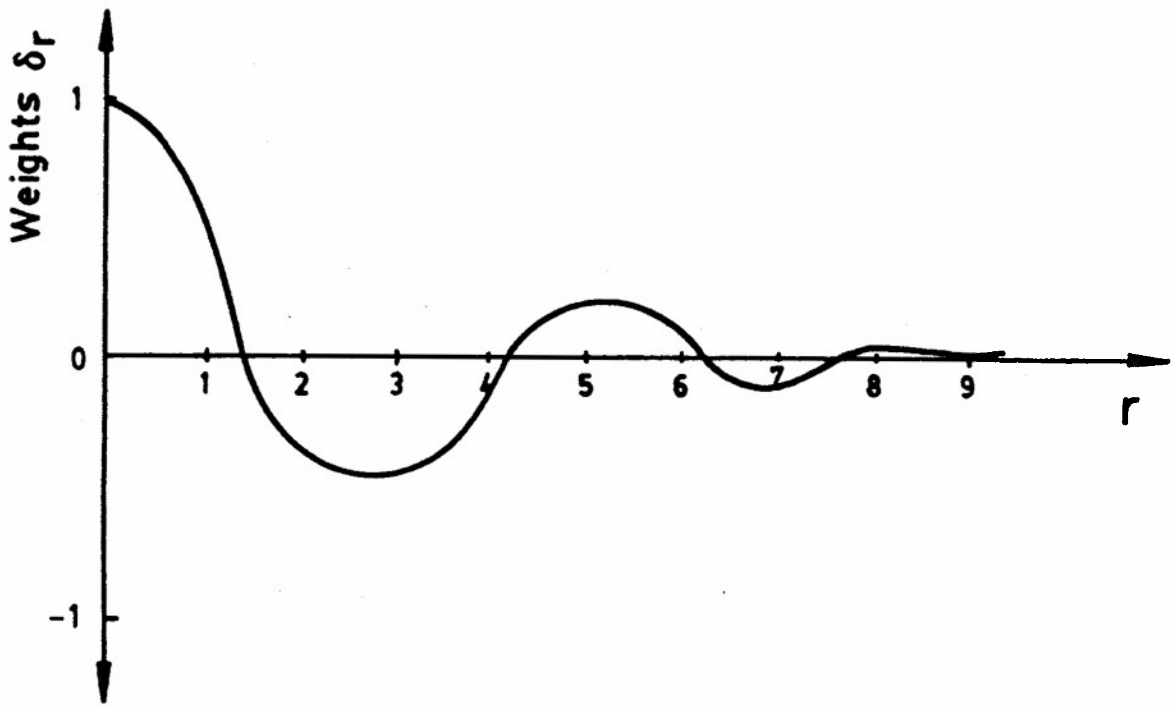

Figure 8.2. Behavior of the weights δ_r as a function of r, for an autoregressive process of order 2 whose associated polynomial equation has complex roots.

In conclusion, we note that for a stationary autoregressive process of order 2, the remote innovations in a (8.15) type representation of the series receive a smaller weight than the more recent ones, regardless of whether the roots of the associated polynomial equation are real or complex. The nature of the roots determines whether the weights decay exponentially or sinusoidally.

8.3 The Covariance Function of an Autoregressive Process

If the joint distributions of the y_t's are normal, then the process is completely determined by its first and second order moments, $\mathcal{E}y_t$, $\mathcal{E}y_t^2$, and $\mathcal{E}y_t y_{t+s}$, $s = 1,2,\ldots,$. If the joint distributions are not normal, the above moments still give us some information about the process. For example, $\mathcal{E}y_t y_{t+s} \big/ \sqrt{\mathcal{E}y_t^2 \mathcal{E}y_s^2}$, the correlation between y_t and y_{t+s} (assuming that $\mathcal{E}y_t = 0$ for all t), is a measure of the relationship between the two variables y_t and y_{t+s} for $t = 1,2,\ldots$.

If the process is stationary, then all the variances are the same, and the covariances depend only on the difference between the two indices. Thus

$$\mathcal{E}y_t y_{t+s} = \sigma(s) = \sigma(-s) , \quad s = \ldots, -1, 0, +1, \ldots .$$

Recall that $\sigma(s)$ is also called the autocovariance function and that $\sigma(s)/\sigma(0)$ is also called the autocorrelation function; it will be denoted by $\rho(s)$ and abbreviated as ACF.

We shall now look at the properties of the covariance function $\sigma(s)$.

If we replace t by $t-s$ in $y_t = \sum_{q=0}^{\infty} \delta_q u_{t-q}$ and multiply it by $\sum_{r=0}^{p} \beta_r y_{t-r} = u_t$, we have

$$(8.19) \qquad \sum_{r=0}^{p} \beta_r y_{t-r} y_{t-s} = \sum_{q=0}^{\infty} \delta_q u_{t-s-q} u_t \ .$$

Now $\mathcal{E} y_{t-r} y_{t-s} = \sigma(s-r)$, $\mathcal{E} u_t^2 = \sigma^2$, $\mathcal{E} u_t u_s = 0$, $t \neq s$, and so the expected value of (8.19) satisfies the following equations:

$$(8.20) \qquad \sum_{r=0}^{p} \beta_r \, \sigma(s-r) = \sigma^2, \qquad s = 0$$

$$(8.21) \qquad \sum_{r=0}^{p} \beta_r \, \sigma(s-r) = 0 \ , \qquad s = 1,2,\ldots, \ .$$

The above equations are known as the Yule-Walker equations; these will be discussed further in Section 8.4.

From (8.21) we observe that the sequence $\sigma(1-p), \sigma(2-p),\ldots, \sigma(0), \sigma(1),\ldots$ satisfies a homogeneous difference equation, which is the same as the homogeneous difference equation (8.12). Thus, if x_1, $\ldots, x_p$, the roots of the polynomial equation $\sum_{r=0}^{p} \beta_r x^{p-r} = 0$, are distinct and $\beta_p \neq 0$, the solution to (8.21) is of the form

$$(8.22) \qquad \sigma(h) = \sum_{i=1}^{p} c_i x_i^h , \qquad h = 1-p, 2-p,\ldots,0,1,\ldots,$$

where $c_1,\ldots,c_p$ are coefficients.

There are $p-1$ boundary conditions of the form

$$\sigma(h) = \sigma(-h)\ , \quad h = 1,\ldots,p-1\ ,$$

and the other boundary condition is given by (8.20) with $\sigma(-p)$ replaced by $\sigma(p)$.

Thus the behavior of the autocovariance function of an AR(p) process is determined by the general nature of (8.22). We study this by considering some special cases.

8.3.1 Special Cases Describing the Behavior of the Autocovariance Function of an Autoregressive Process

Following Section 8.2.1 , we consider here an autoregressive process of order 1 and an autoregressive process of order 2.

An Autoregressive Process of Order 1

Suppose that in (8.1) $p = 1$ (and $\mu = 0$), so that

$$y_t + \beta_1 y_{t-1} = u_t\ , \qquad t = 2,3,\ldots,\ .$$

The associated polynomial equation $x + \beta_1 x^0 = 0$ has one root $x_1 = -\beta_1$. The general solution is (from (8.22)) $\sigma(h) = c_1(-\beta_1)^h$, $h = 0,1,\ldots$. From (8.20) we have

$$\sigma^2 = \sigma(0) + \beta_1\sigma(1) = c_1[1 + \beta_1(-\beta_1)] = c_1[1 - \beta_1^2]\quad .$$

Hence $c_1 = \sigma^2/(1-\beta_1^2)$, so that

$$\sigma(h) = (-\beta_1)^h \sigma^2/(1-\beta_1^2) , \qquad h=0,1,\ldots .$$

From $\rho(h) = \sigma(h)/\sigma(0)$, the autocorrelation function is

$$(8.23) \qquad \rho(h) = (-\beta_1)^h , \qquad h=0,1,\ldots .$$

If $|\beta_1| < 1$, then we have the important useful result that the theoretical autocorrelation function of an autoregressive process of order 1 decays exponentially in the lag h . The decay is smooth if $\beta_1 < 0$, and it alternates in sign if $\beta_1 > 0$. In Figure 8.3, we illustrate this behavior of $\rho(h)$ for nonnegative values of h . We also remark that the behavior of $\rho(h)$ is analogous to the behavior of the weights δ_r discussed in Section 8.2.1 - see (8.14) and Figure 8.2.

An Autoregressive Process of Order 2

Now suppose that in (8.1) $p=2$ (and $\mu=0$) , so that

$$y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} = u_t , \qquad t=3,4,\ldots, .$$

The associated polynomial equation $x^2 + \beta_1 x + \beta_2 x^0 = 0$ has the roots $x_i = [-\beta_1 \pm \sqrt{\beta_1^2 - 4\beta_2}]/2$, $i=1,2$.

If the roots x_1 and x_2 are distinct, then $\sigma(h) = c_1 x_1^h + c_2 x_2^h$, $h=-1,0,1,\ldots$. Then (8.20) and $\sigma(1) = \sigma(-1)$ can be solved for c_1 and c_2 , yielding

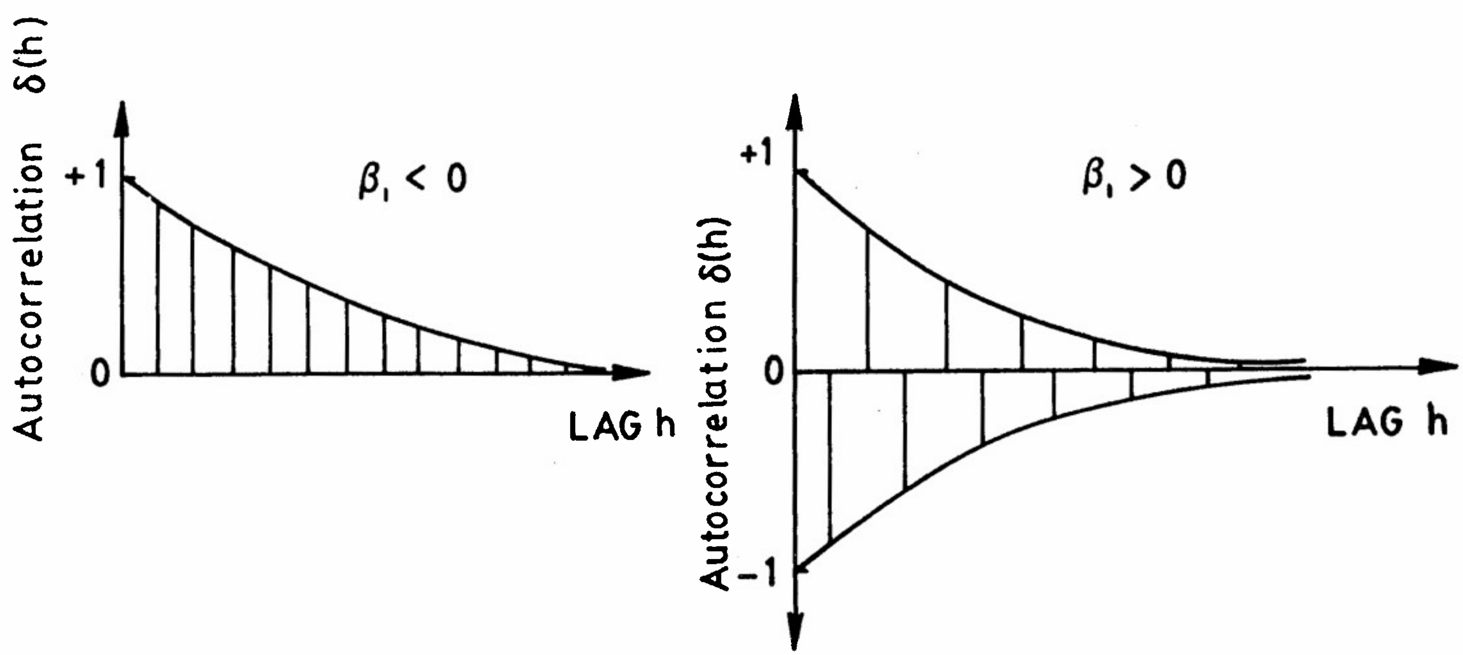

Figure 8.3. Behavior of the theoretical autocorrelation function of an autoregressive process of order 1.

$$(8.24) \qquad \sigma(h) = \frac{\sigma^2}{(x_1-x_2)(1-x_1x_2)} \left(\frac{x_1^{h+1}}{1-x_1^2} - \frac{x_2^{h+1}}{1-x_2^2}\right) , \qquad h = 0,1,\ldots \quad .$$

If we require that $|x_i| < 1$, $i=1,2,$ then β_1 and β_2 must lie in the triangular region described by Figure 8.1; that is, they must satisfy the inequalities (8.17). Furthermore, if x_1 and x_2 are real, that is β_1 and β_2 do not lie in the parabolic region of Figure 8.1, so that $\beta_1^2 > 4\beta_2$, then by (8.24) we have the result that $\sigma(h)$ is a linear combination of two exponentially decaying functions of h , x_1^{h+1} and x_2^{h+1} . Depending on whether the dominant root is positive or negative, $\sigma(h)$ will remain positive or alternate in sign as it damps out. This behavior of $\sigma(h)$ as a function of $h \geq 0$, is shown in Figure 8.4.

When $|x_i| < 1$, $i = 1,2,$ and x_1 and x_2 are complex, that is, $\beta_1^2 < 4\beta_2$, then x_1 and x_2 can be written as $x_1 = \alpha e^{i\theta}$ and $x_2 = \alpha e^{-i\theta}$, where $\alpha < 1$, and now (8.24) becomes

$$(8.25) \qquad \sigma(h) = \frac{\sigma^2\alpha^h[\sin\theta(h+1) - \alpha^2 \sin\theta(h-1)]}{(1-\alpha^2)\sin\theta[1-2\alpha^2 \cos 2\theta+\alpha^4]} ,$$

$$= \frac{\sigma^2\alpha^h \cos(\theta h-\phi)}{(1-\alpha^2)\sin\theta\sqrt{1-2\alpha^2 \cos 2\theta+\alpha^4}} , \qquad h = 0,1,\ldots,$$

where $\tan\phi = (1-\alpha^2)\cot\theta/(1+\alpha^2)$.

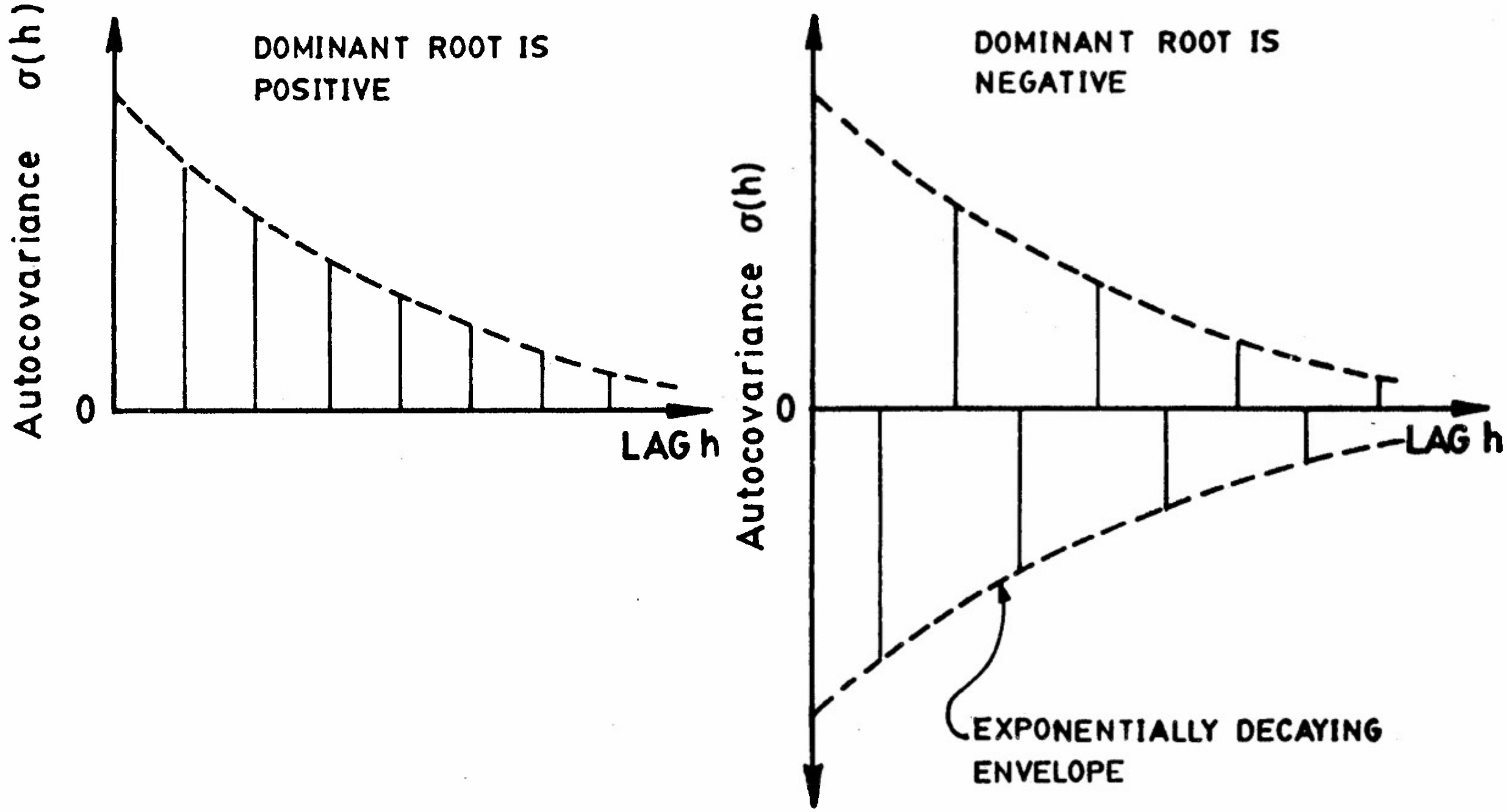

Figure 8.4. Behavior of $\sigma(h)$ the theoretical autocovariance function of an autoregressive process of order 2, when its associated polynomial equation has real roots.

Thus $\sigma(h)$ is a damped cosine function of h; the behavior of $\sigma(h)$ as a function of $h = 0, \pm 1, \pm 2, \ldots,$ is illustrated in Figure 8.5.

Since $\sigma(h)$ is a linear combination of the hth powers of the roots x_1 and x_2, both of which are less than 1 in absolute value, $|\sigma(h)|$ is bounded. We remark that the behavior of $\sigma(h)$ as a function of h is analogous to the behavior of the weights δ_r as a function of r, discussed in Section 8.2.1 - see (8.16), (8.18), and Figure 8.2.

Thus to conclude, we have the important practical result, that when β_1 and β_2, the parameters of an AR(2) process, lie in the triangular region described by Figure 8.1, the theoretical autocorrelation function decays either exponentially or sinusoidally. The exponential decay could be either smooth or alternating in sign, depending on the values that β_1 and β_2 take.

In Section 8.11, we show the behavior of the estimated autocorrelation function of some real life data which we claim can be reasonably well described by autoregressive processes. However, in order to be able to use the behavior of the autocorrelation function as a means of identifying autoregressive processes, we need to have some idea about the behavior of the estimated autocorrelation function of some <u>known</u> autoregressive processes. This we do next, and also make some other comments which have some practical implications.

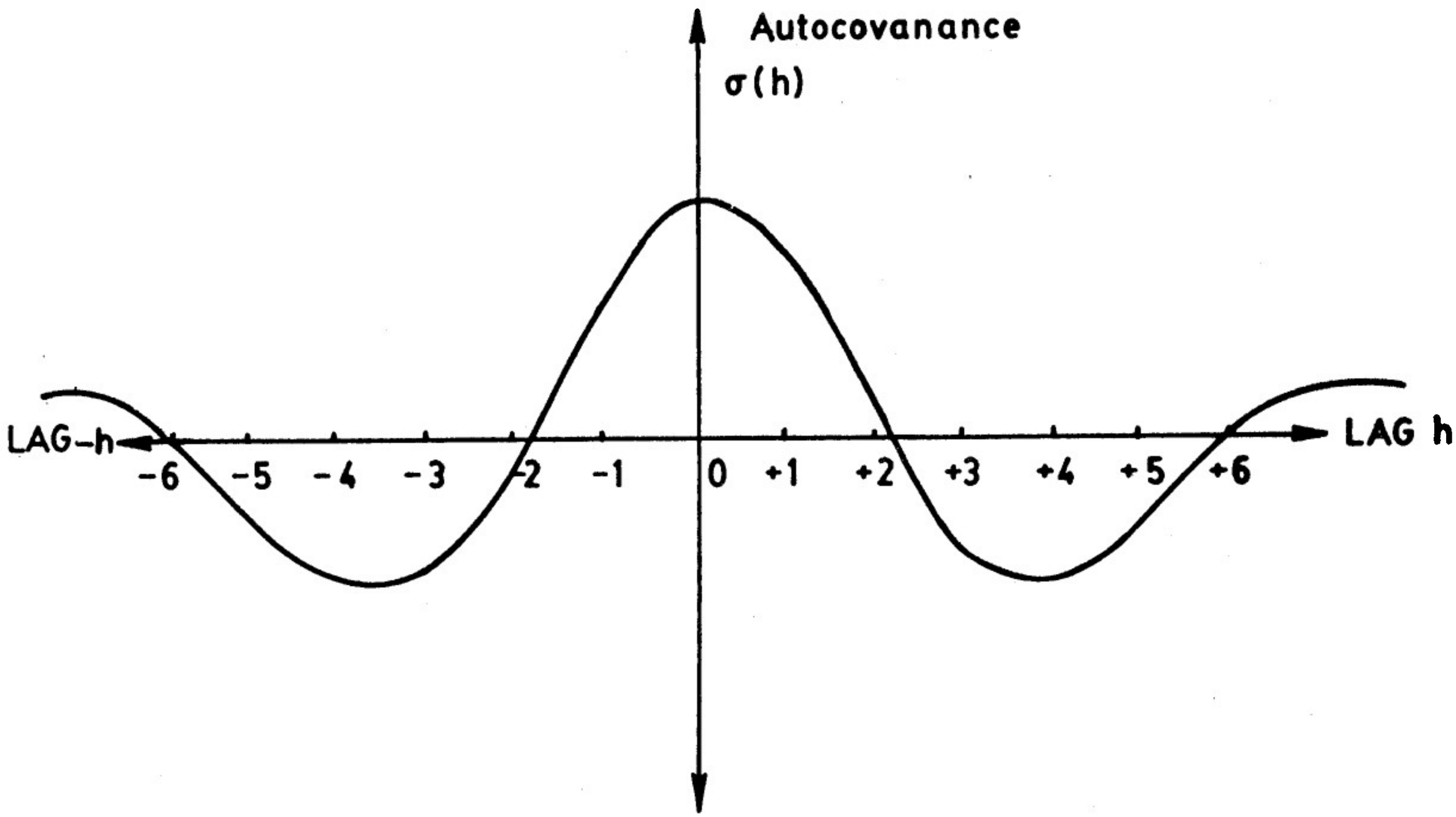

Figure 8.5. Behavior of $\sigma(h)$ the theoretical autocovariance function of an autoregressive process of order 2 when its associated polynomial equation has complex roots.

8.3.2 Behavior of the Estimated Autocorrelation Function of Some Simulated Autoregressive Processes

The results of Section 8.3.1 can be generalized in a straightforward manner to show that the autocorrelation function of autoregressive processes must decay exponentially or sinusoidally. Even though this result is true in theory, it is unreasonable to expect such a behavior of the estimated autocorrelation function. Such a lack of conformance between the theory and its application is mainly due to the sampling variability in our estimate of the autocorrelation function (see Section 7), and is particularly acute when we are dealing with series of short lengths, wherein our estimate of the autocorrelation function is based on few observations. Thus a good deal of caution and insight has to be used in order to identify the nature of an underlying stochastic process by examining the behavior of its estimated autocorrelation function.

In Table 8.1 we give $r(h)$, the values of the estimated autocorrelation function, $h=0,1,\ldots,25$, based on 250 computer generated observations from an AR(1) process

$$y_t - .5y_{t-1} = u_t \; , \qquad t=2,3,\ldots,250,$$

with $y_1 = u_1$. A plot of $r(h)$ versus h is given in Figure 8.6. Barring the slight aberrations at $h=7$, 8, 9, 13, 19, and 23, this plot reveals the exponential decay pattern expected of an AR(1) process with $\beta_1 < 0$, and $|\beta_1| < 1$.

In Table 8.2 we give $r(h)$, the values of the estimated autocorrelation function, $h=0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process

Table 8.1

Values of the estimated autocorrelation function $r(h)$, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(1) process with $\beta_1 = -.5$.

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	1.0	.56	.33	.16	.09	.04	.04	.10	.11	.11	.06	.05	.004
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of r(h)	.083	.037	.030	.022	.037	.039	.111	.06	.05	.04	.07	.06	.04

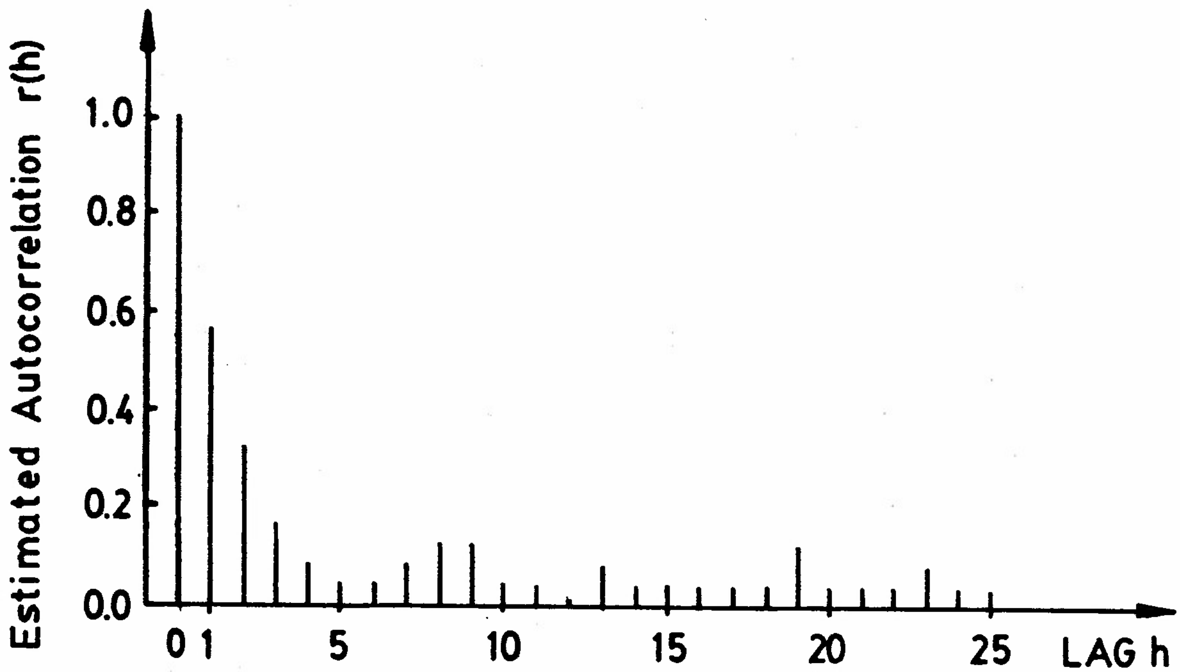

Figure 8.6. A plot of the estimated autocorrelation function $r(h)$ versus h, $h=0,1,\ldots,25$, based on 250 computer generated observations from an AR(1) process with $\beta_1 = -.5$.

Table 8.2

Values of the estimated autocorrelation function $r(h)$, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = -.9$, and $\beta_2 = .4$

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	1.0	.68	.24	-.07	-.16	-.13	-.03	.088	.15	.13	.06	.01	-.01
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of r(h)	.03	.03	.01	.00	.02	.05	.09	.05	.02	.01	.04	.05	.06

$$y_t - .9y_{t-1} + .4y_{t-2} = u_t \ , \qquad t = 3,4,\ldots,250,$$

with $y_2 = .9y_1 + u_2$ and $y_1 = u_1$. A plot of $r(h)$ versus h is given in Figure 8.7. Since $\beta_1 = -.9$ and $\beta_2 = .4$, $\beta_1^2 < 4\beta_2$, and so the roots of the associated polynomial equation are complex. ($\alpha = \sqrt{.4} = .63$, $x_1, x_2 = .45 \pm .44i$, $\theta \sim 45^\circ$) Thus the theoretical autocorrelation function must decay sinusoidally; this feature is also revealed by the estimated autocorrelation function shown in Figure 8.7.

In Table 8.3 we give $r(h)$, the values of the estimated autocorrelation function, for $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process

$$y_t + .5y_{t-1} - .2y_{t-2} = u_t \ , \qquad t = 3,4,\ldots, \quad .$$

A plot of $r(h)$ versus h is given in Figure 8.8. Since $\beta_1 = .5$, and $\beta_2 = -.2$, $\beta_1^2 > 4\beta_2$, and hence the roots of the associated polynomial equation are real. These roots being $(-.5 \pm \sqrt{.25+.8})/2$, it is clear that the dominant root is negative, its value is $\frac{-.5 - 1.025}{2} = -.763$. Thus according to the material in Section 8.3.1, the autocorrelation function must decay, and alternate in sign as it does so - see Figure 8.4. The estimated autocorrelation function of Figure 8.8 reveals this tendency, at least in the earlier stages, up to lag 10 or so. Later on, the estimated autocorrelation function does alternate in sign, but does not decay. We attribute our reasons for this to the sampling variability of the estimates of the autocorrelation at the various lags.

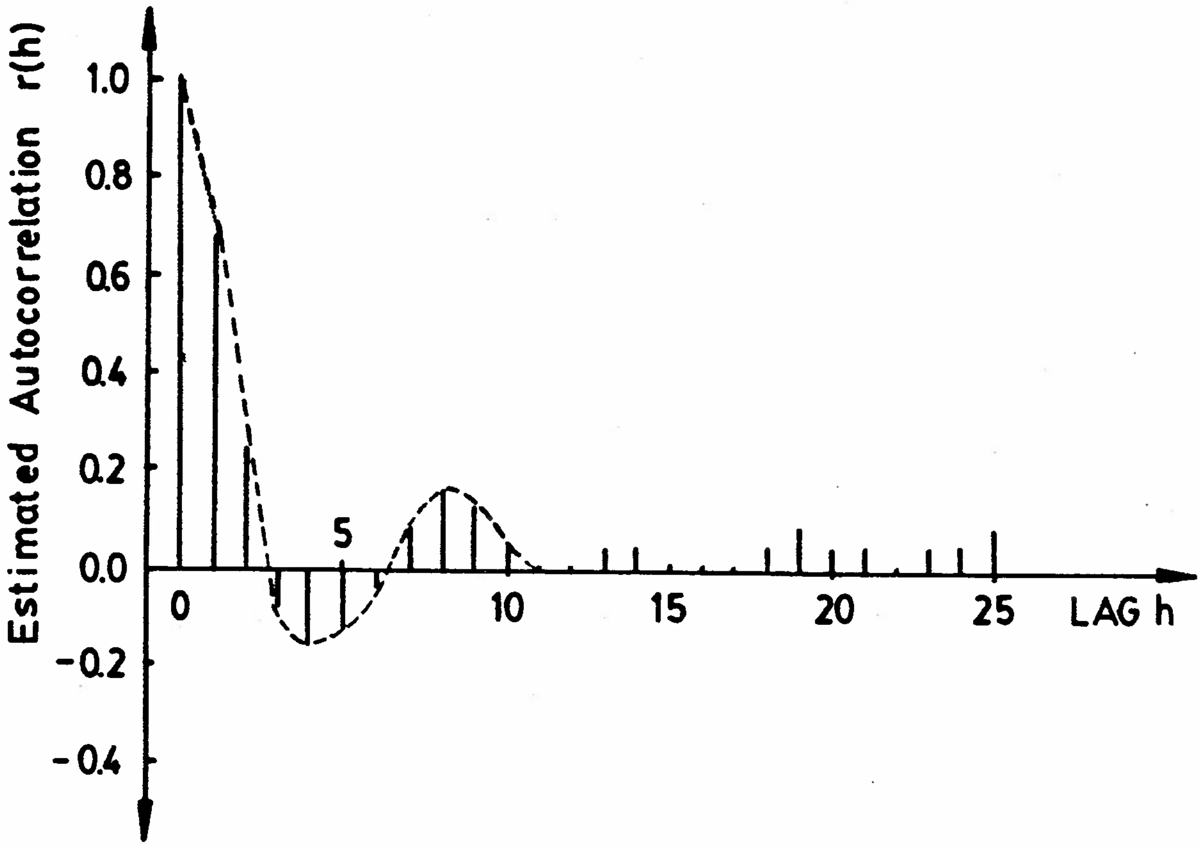

Figure 8.7. A plot of the estimated autocorrelation function $r(h)$ versus h, $h = 0,1,\ldots,25$, based on 250 computer generated obervations from an AR(2) process with $\beta_1 = -.9$ and $\beta_2 = .4$.

Table 8.3

Values of the estimated autocorrelation function $r(h)$, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = .5$ and $\beta_2 = -.2$

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	1	-.63	.56	-.40	.30	-.19	.09	-.01	-.01	.10	-.12	.16	-.20
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of r(h)	.22	-.19	.17	-.17	.17	-.19	.22	-.17	.15	-.12	.10	-.04	-.01

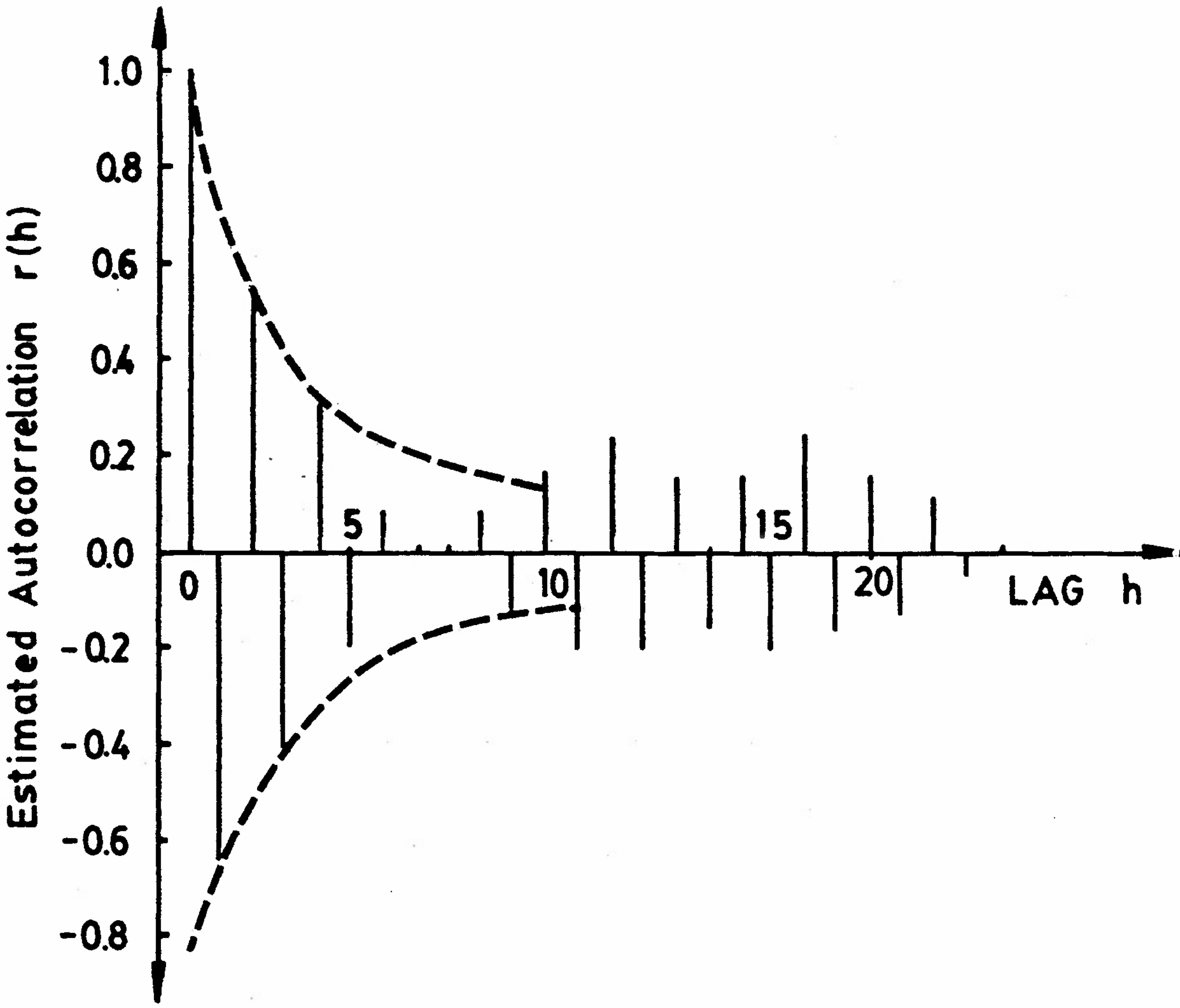

Figure 8.8. A plot of the estimated autocorrelation function $r(h)$ versus h, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process, with $\beta_1 = .5$ and $\beta_2 = -.2$.

The behavior of the estimated autocorrelation function of some real life data which we feel can be reasonably well described by autoregressive processes is shown at the end of this section, in 8.11.

8.4 Expressing the Parameters of an Autoregressive Process in Terms of its Autocorrelations

The Yule-Walker equations (8.21) enable us to express the autoregressive parameters $\beta_1,\ldots,\beta_p$ in terms of the autocorrelations $\rho(s)$, $s=1,\ldots,$. To see this, we set $s=1,\ldots,p$ in (8.21), divide throughout by $\sigma(0)$, and observe that

$$
\text{(8.26)} \qquad \begin{cases} \rho(1) = -\beta_1 - \beta_2\rho(1) - \cdots - \beta_p\rho(p-1) \ , \\ \rho(2) = -\beta_1\rho(1) - \beta_2 - \cdots - \beta_p\rho(p-2) \ , \\ \vdots \qquad\qquad\qquad\qquad \vdots \\ \rho(p) = -\beta_1\rho(p-1)-\beta_2\rho(p-2) - \cdots - \beta_p \ . \end{cases}
$$

If we denote $\underset{\sim}{\beta} = [\beta_1,\ldots,\beta_p]'$, $\underset{\sim}{\rho} = [\rho(1),\ldots,\rho(p)]'$, and

$$
\underset{\sim}{\mathbb{P}} = \begin{bmatrix} 1 & \rho(1) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \cdots & \rho(p-2) \\ \vdots & & & \vdots \\ \rho(p-1) & \rho(p-2) & \cdots & 1 \end{bmatrix}
$$

then $\underset{\sim}{\rho} = -\underset{\sim}{\mathbb{P}}\underset{\sim}{\beta}$ from which we have (since $\underset{\sim}{\mathbb{P}}$ is positive definite)

$$
\text{(8.27)} \qquad \underset{\sim}{\beta} = -\underset{\sim}{\mathbb{P}}^{-1}\underset{\sim}{\rho} \ .
$$

The matrix $\underset{\sim}{\mathbb{P}}$ is unknown as the <u>autocorrelation matrix</u>.

Thus, the p autoregressive parameters can be expressed in terms of the p autocorrelations $\rho(1),\ldots,\rho(p)$. This feature can be used to estimate $\underset{\sim}{\beta}$, using an estimate of $\underset{\sim}{\mathbb{P}}$.

We obtain σ^2, the variance of the disturbance, by setting $\sigma(-r) = \sigma(r)$ in the Yule-Walker equation (8.20) to obtain

$$\beta_0\sigma(0) + \beta_1\sigma(1) + \ldots + \beta_p\sigma(p) = \sigma^2 \quad . \tag{8.28}$$

8.5 The Partial Autocorrelation Function of an Autoregressive Process

In Section 8.3 we have shown that $\sigma(h)$, the autocovariance function of an autoregressive process of order p, is infinite in extent. Thus from $\{\sigma(h)\}$ it is hard to determine the order of an autoregressive process. The partial autocorrelation function, to be discussed here, will help us in determing the order of an autoregressive process.

To be specific, let us consider a stationary autoregressive process of order p

$$y_t = u_t - \beta_1 y_{t-1} - \ldots - \beta_p y_{t-p}\,, \qquad t = p+1,\ p+2,\ldots \quad .$$

Recall that in order to predict y_t we need consider only the p lagged variables $y_{t-1},\ldots,y_{t-p}$, since the other variables y_{t-p-1}, $y_{t-p-2},\ldots$ have no effect on y_t.

The partial autocorrelation between y_t and y_{t-p}, to be denoted by $\pi(p)$ is the correlation between y_t and y_{t-p} when the intermediate $p-1$ variables $y_{t-1}, y_{t-2},\ldots,y_{t-p+1}$ are "held fixed." That is, $\pi(p)$ is the correlation between y_t and y_{t-p} when the intermediate variables are not allowed to vary and exert their influence on the relationship between y_t and y_{t-p}. Clearly, $\pi(1)$, the partial autocorrelation between y_t and y_{t-1}, is $\rho(1)$, the (ordinary) autocorrelation between y_t and y_{t-1}, whereas $\pi(0)$ the partial autocorrelation between y_t and itself is 1.

Thus, by its very nature, since $y_{t-p-1}, y_{t-p-2},\ldots,$ have no effect on y_t, the partial autocorrelation function of an autoregressive process of order p, $\pi(j) \neq 0$, for $j=0,1,\ldots,p$, and $\pi(j)=0$, for $j>p$. The fact $\pi(j)$ vanished for $j \geq p+1$, can be used to identify the order p of an autoregressive process, provided that $\pi(j)$ can be computed.

In our discussion of the partial autocorrelation function $\pi(p)$ we had mentioned the fact that the intermediate values $y_{t-1},\ldots,y_{t-p+1}$ had to be "held fixed". In order to formalize this notion we shall use some results which are standard in multivariate analysis.

Let $\underset{\sim}{Y} = [y_t, y_{t-1},\ldots,y_{t-p}]$ denote the vector of $p+1$ observations, and let $\underset{\sim}{\Sigma}$ denote the variance-covariance matrix of these $p+1$ observations. Suppose that $\underset{\sim}{Y}$ has a multivariate normal distribution with mean vector $\underset{\sim}{0}$ and covariance matrix $\underset{\sim}{\Sigma}$, where

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma(0) & \sigma(1) & \cdots & \sigma(p) \\ \sigma(1) & \sigma(0) & \cdots & \sigma(p-1) \\ \vdots & \vdots & & \vdots \\ \sigma(p) & \sigma(p-1) & \cdots & \sigma(0) \end{bmatrix} .$$

Let us rearrange the elements of $\underset{\sim}{Y}$, and partition it into two component sub-vectors $\underset{\sim}{Y}^{(1)} = [y_t, y_{t-p}]$ and $\underset{\sim}{Y}^{(2)} = [y_{t-1}, y_{t-2}, \ldots, y_{t-p+1}]$. Let $\underset{\sim}{\Sigma}_{11}$, $\underset{\sim}{\Sigma}_{22}$, and $\underset{\sim}{\Sigma}_{12}$ be the variance-covariance matrices of $\underset{\sim}{Y}^{(1)}$, $\underset{\sim}{Y}^{(2)}$, and $\underset{\sim}{Y}^{(1)}$ and $\underset{\sim}{Y}^{(2)}$ respectively. That is, $\underset{\sim}{\Sigma}_{11}$, $\underset{\sim}{\Sigma}_{22}$, and $\underset{\sim}{\Sigma}_{12}$ is a partition of the rearrangement of $\underset{\sim}{\Sigma}$.

Let $\underset{\sim}{y}^{(2)}$ be a particular value taken by the vector $\underset{\sim}{Y}^{(2)}$. Then, it can be shown [Anderson (1984), p.28] that the conditional distribution of $\underset{\sim}{Y}^{(1)}$ given $\underset{\sim}{y}^{(2)}$ is a multivariate normal with mean $\underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{y}^{(2)}$, and covariance matrix $\underset{\sim}{\Sigma}_{11} - \underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{\Sigma}_{12}^{'} = \underset{\sim}{\Sigma}_{11\cdot 2}$, say. This is a generalization of the results mentioned in Section 6.

The vector $\underset{\sim}{\Sigma}_{12}\underset{\sim}{\Sigma}_{22}^{-1}\underset{\sim}{y}^{(2)}$ is called the <u>regression function</u> of the regression of $\underset{\sim}{Y}^{(1)}$ on $\underset{\sim}{y}^{(2)}$. The matrix $\underset{\sim}{\Sigma}_{11\cdot 2}$ is a 2×2 matrix whose elements are indicated below:

$$\underset{\sim}{\Sigma}_{11\cdot 2} = \begin{bmatrix} \sigma_{tt\cdot(t-1),\ldots,(t-p+1)} & \sigma_{t(t-p)\cdot(t-1),\ldots,(t-p+1)} \\ \sigma_{(t-p)t\cdot(t-1),\ldots,(t-p+1)} & \sigma_{(t-p)(t-p)\cdot(t-1),\ldots,(t-p+1)} \end{bmatrix} .$$

The <u>partial correlation</u> between y_t and y_{t-p} holding $(t-1),\ldots,(t-p+1)$ fixed at $\underset{\sim}{y}^{(2)}$ is

$$\pi(p) = \frac{\sigma_{t(t-p)\cdot(t-1),\ldots,(t-p+1)}}{\sqrt{\sigma_{tt\cdot(t-1)\ldots(t-p+1)}}\ \sqrt{\sigma_{(t-p)(t-p)\cdot(t-1),\ldots,(t-p+1)}}} \quad ;$$

note that $\pi(p)$ is independent of $\underset{\sim}{y}^{(2)}$.

As an example, if $\underset{\sim}{Y} = (y_t, y_{t-1}, y_{t-2})'$, and if $\underset{\sim}{Y}^{(1)} = (y_t, y_{t-2})'$ and $\underset{\sim}{Y}^{(2)} = y_{t-1}$, then the partial correlation between y_t and y_{t-2}, $\pi(2)$, turns out to be

$$\pi(2) = (\rho(2) - \rho^2(1))/(1 - \rho^2(1)) \ .$$

8.5.1 Relationship between Partial Autocorrelation and the Last Coefficient of an Autoregressive Process

An interesting relationship between $\pi(p)$, the partial autocorrelation of y_t and y_{t-p}, and β_p, the last coefficient of an autoregressive process of order p, can be observed. This relationship simplifies our calculation of $\pi(p)$, since β_p can be easily obtained from the Yule-Walker equations via equation (8.28).

In order to see a relationship between $\pi(p)$ and β_p, consider an AR(2) process

$$y_t = u_t - \beta_1 y_{t-1} - \beta_2 y_{t-2}$$

and solve the resulting Yule-Walker equations to obtain

$$\beta_2 = - \frac{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & \rho(2) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) \\ \rho(1) & 1 \end{vmatrix}} = - \frac{\rho(2) - \rho^2(1)}{1 - \rho^2(1)} .$$

However $[(\rho(2) - \rho^2(1))/(1 - \rho^2(1))]$ is indeed the partial autocorrelation between y_t and y_{t-2}; thus $\pi(2) = -\beta_2$. In a similar manner, if we consider an AR(3) process

$$y_t = u_t - \beta_1 y_{t-1} - \beta_2 y_{t-2} - \beta_3 y_{t-3}$$

and solve the resulting Yule-Walker equations, we observe that

$$\beta_3 = - \frac{\begin{vmatrix} 1 & \rho(1) & \rho(1) \\ \rho(1) & 1 & \rho(2) \\ \rho(2) & \rho(1) & \rho(3) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) \\ \rho(1) & 1 & \rho(1) \\ \rho(2) & \rho(1) & 1 \end{vmatrix}} ,$$

which again can be verified as the negative of the partial autocorrelation between y_t and y_{t-3}.

In general, we observe [Anderson (1971), pages 188 and 222] that for an autoregressive process or order p, $\pi(p)$ the partial autocorrelation between y_t and y_{t-p} is $-\beta_p$, where

$$(8.29) \quad \beta_p = - \frac{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(2) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & \rho(p) \end{vmatrix}}{\begin{vmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(p-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(p-2) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(p-1) & \rho(p-2) & \rho(p-3) & \cdots & 1 \end{vmatrix}} .$$

It is helpful to remark that the determinant in the denominator is simply the determinant of the autocorrelation matrix for an AR(p) process $\underset{\sim}{\mathbb{P}}$ (Section 8.4), whereas the matrix in the numerator is $\underset{\sim}{\mathbb{P}}$ with the last column replaced by $\rho(1),\ldots,\rho(p)$.

An expression for $\pi(j)$ the partial autocorrelation between y_t and y_{t-j} , can be obtained if we write the Yule-Walker equations for j , and set $\pi(j) = \beta_j$, where β_j is given by equation (8.29); recall that $\pi(0) = 1$, and that $\pi(1) = \rho(1)$.

The <u>partial autocorrelation function</u> is a plot of $\pi(h)$ versus h , $h = 1,2,\ldots,$; the partial autocorrelation function is abbreviated as PACF.

We estimate $\pi(1)$ by $r(1)$, and estimate $\pi(j)$ by $\hat{\pi}(j)$, where $\hat{\pi}(j)$ is obtained by replacing the $\rho(\cdot)$'s in (8.29) by their estimates $r(\cdot)$'s .

8.5.2 Behavior of the Estimated Partial Autocorrelation Function of Some Simulated Autoregressive Processes

Even though the partial autocorrelation function of an autoregressive process of order p must theoretically vanish at lags $p+1$, $P+2$, ..., it is unreasonable to expect such a behavior of the estimated partial autocorrelation function. The reasons for this are analogous to those given for the behavior of the estimated autocorrelation function - see Section 8.3.2. Thus caution and insight must be used when identifying the order of an autoregressive process by examining its estimated partial autocorrelation function.

In Table 8.4 we give $\hat{\pi}(h)$, the values of the partial autocorrelation function for $h=0,1,\ldots,25$ based on 250 computer generated observations from the AR(1) process

$$y_t - .5y_{t-1} = u_t \ , \qquad t=2,3,\ldots \ ,$$

discussed in Seciton 8.3.2.

A plot of $\hat{\pi}(h)$ versus h is given in Figure 8.9. Barring some slight aberrations at a few lags, this plot reveals the behavior that we expect from the PACF of an AR(1) process, namely that $\hat{\pi}(1)$ must be significantly different from 0, and that $\hat{\pi}(j)$, must be close to 0 for $j=2,3,\ldots,$.

An examination of Figures 8.6 and 8.9 reveals the desired result that for an AR(1) process, the autocorrelation function decays exponentially, and that the partial autocorrelation function vanishes after lag 1.

Table 8.4

Values of the estimated partial autocorrelation function $\hat{\pi}(h)$, $h = 0,1,\ldots,25$, based on computer generated observations from an AR(1) process with $\beta_1 = -.5$

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of $\hat{\pi}(h)$	1	.56	.03	-.06	.03	-.01	.02	.11	.02	.02	-.05	.02	-.03
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of $\hat{\pi}(h)$	.13	-.08	-.01	.01	.02	.01	.14	-.11	.03	.00	.08	-.03	.03

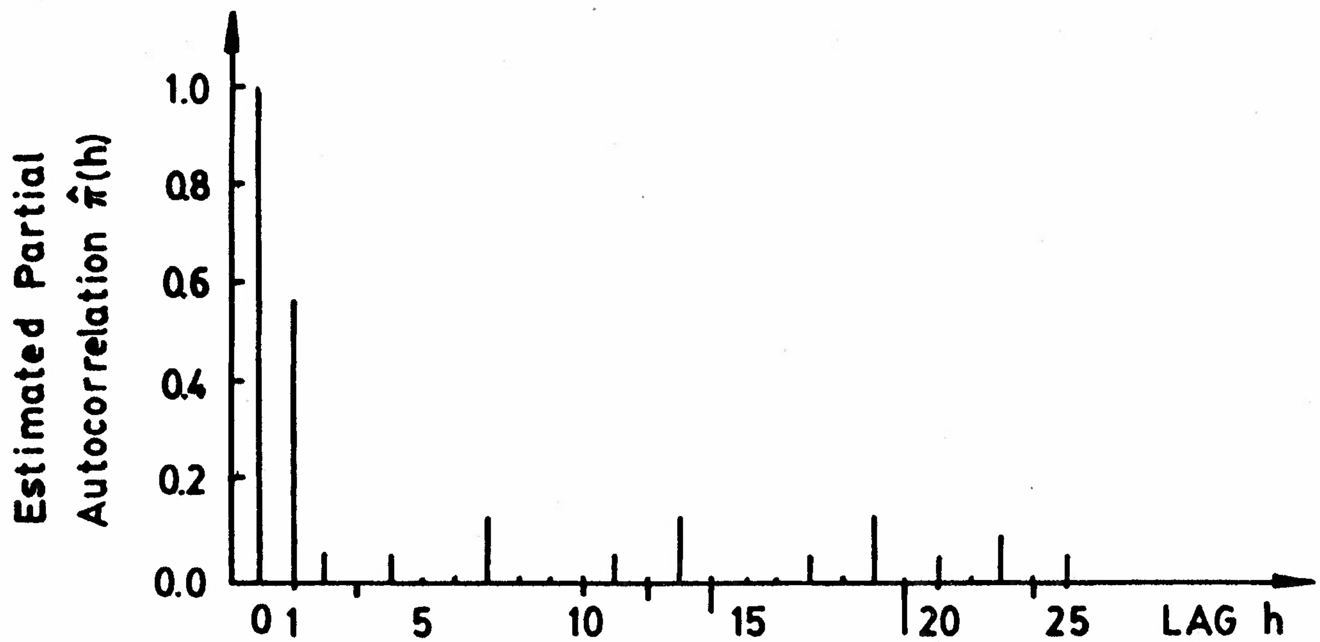

Figure 8.9. A plot of the estimated partial autocorrelation function $\hat{\pi}(h)$ versus h, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = -.5$.

In Table 8.5 we give $\hat{\pi}(h)$, the values of the estimated partial autocorrelation function, for $h=0,1,\ldots,25$, based on 250 computer generated observations from the AR(2) process

$$y_t - .9y_{t-1} + .4y_{t-2} = u_t , \qquad t=3,4,\ldots, ,$$

discussed in Section 8.3.2.

A plot of $\hat{\pi}(h)$ versus h is given in Figure 8.10. As is to be expected, barring some minor aberrations, $\hat{\pi}(1)$ and $\hat{\pi}(2)$ are significantly different from 0, and $\hat{\pi}(j)$ is close to zero, for $j=3,4,\ldots,$.

In Table 8.6 we give $\hat{\pi}(h)$, the values of the estimated partial autocorrelation function, for $h=0,1,\ldots,25$, based on 250 computer generated observations from the AR(2) process

$$y_t + .5y_{t-1} - .2y_{t-2} = u_t , \qquad t=3,4,\ldots ,$$

discussed in Section 8.3.2. A plot of $\hat{\pi}(h)$ versus h is given in Figure 8.11. Once again, as is to be expected, $\hat{\pi}(1)$ and $\hat{\pi}(2)$ are significantly different from 0 , and $\hat{\pi}(j)$ is close to zero, for $j=3,4,\ldots,$.

An inspection of Figures 8.7 and 8.10, and Figures 8.8 and 8.11, reveals the desired result that the autocorrelation function of an AR(2) process decays either exponentially or sinusoidally, whereas the partial autocorrelation function vanishes after lag 2.

The behavior of the estimated autocorrelation function and the partial autocorrelation function of some real life data which we believe can be reasonably well approximated by autoregressive processes is shown in Section 8.11.

Table 8.5

Values of the estimated partial autocorrelation function $\hat{\pi}(h)$, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = -.9$ and $\beta_2 = .4$.

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of $\hat{\pi}(h)$	1	.68	-.41	-.04	.03	-.02	.06	.09	-.01	.02	-.03	.05	.01
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of $\hat{\pi}(h)$	.09	-.12	.04	.01	.03	.03	.07	-.13	.09	.01	.07	-.03	.05

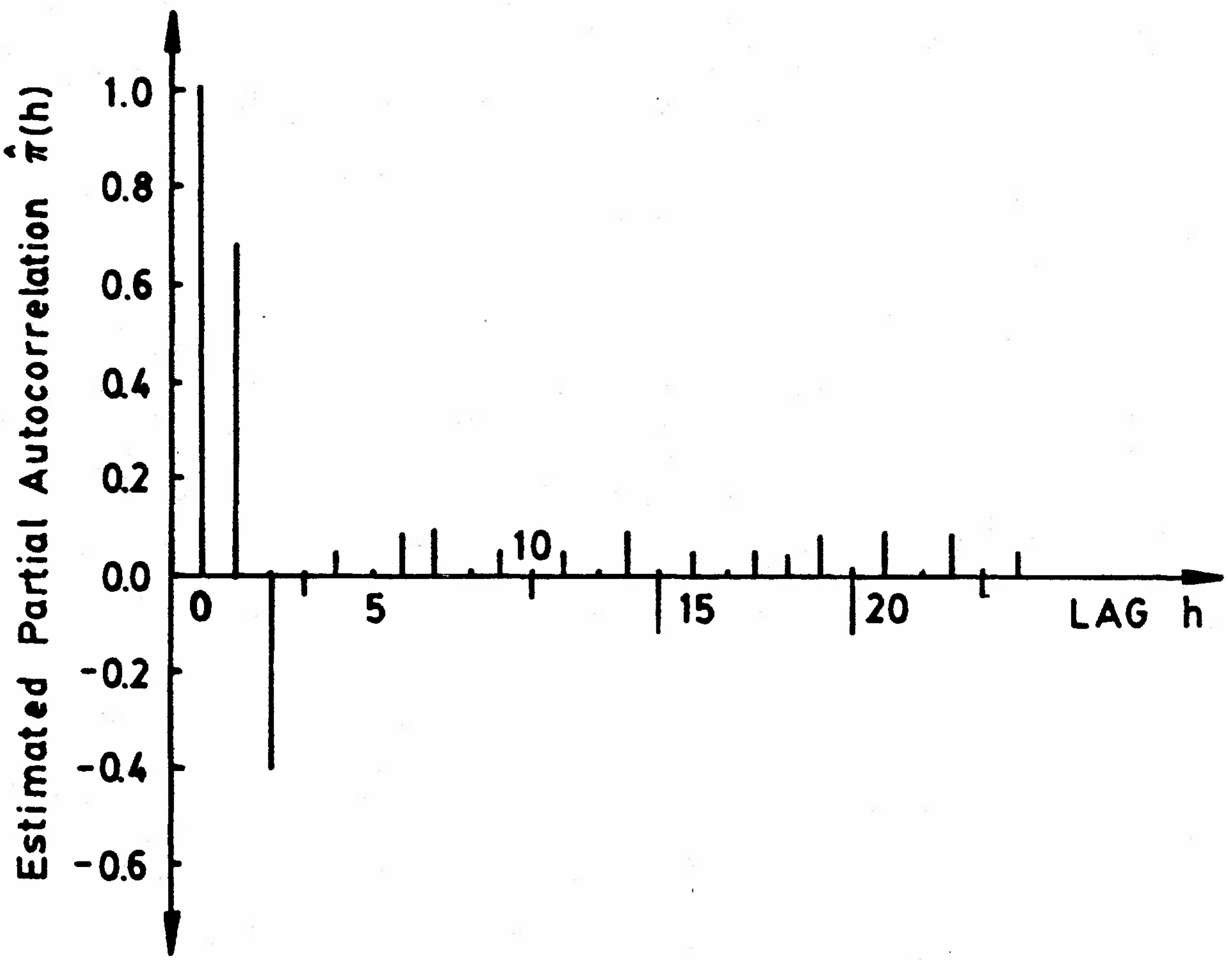

Figure 8.10. A plot of the estimated partial autocorrelation function $\hat{\pi}(h)$ versus h, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = -.9$ and $\beta_2 = .4$.

Table 8.6

Values of the estimated partial autocorrelation function $\hat{\pi}(h)$, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = .5$ and $\beta_2 = -.2$.

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of $\hat{\pi}(h)$	1	-.63	.27	.03	-.05	.06	-.07	.06	.05	.10	-.05	.04	-.06
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of $\hat{\pi}(h)$	.06	.05	-.01	-.06	.06	-.08	.12	.07	-.03	-.02	.03	.05	-.01

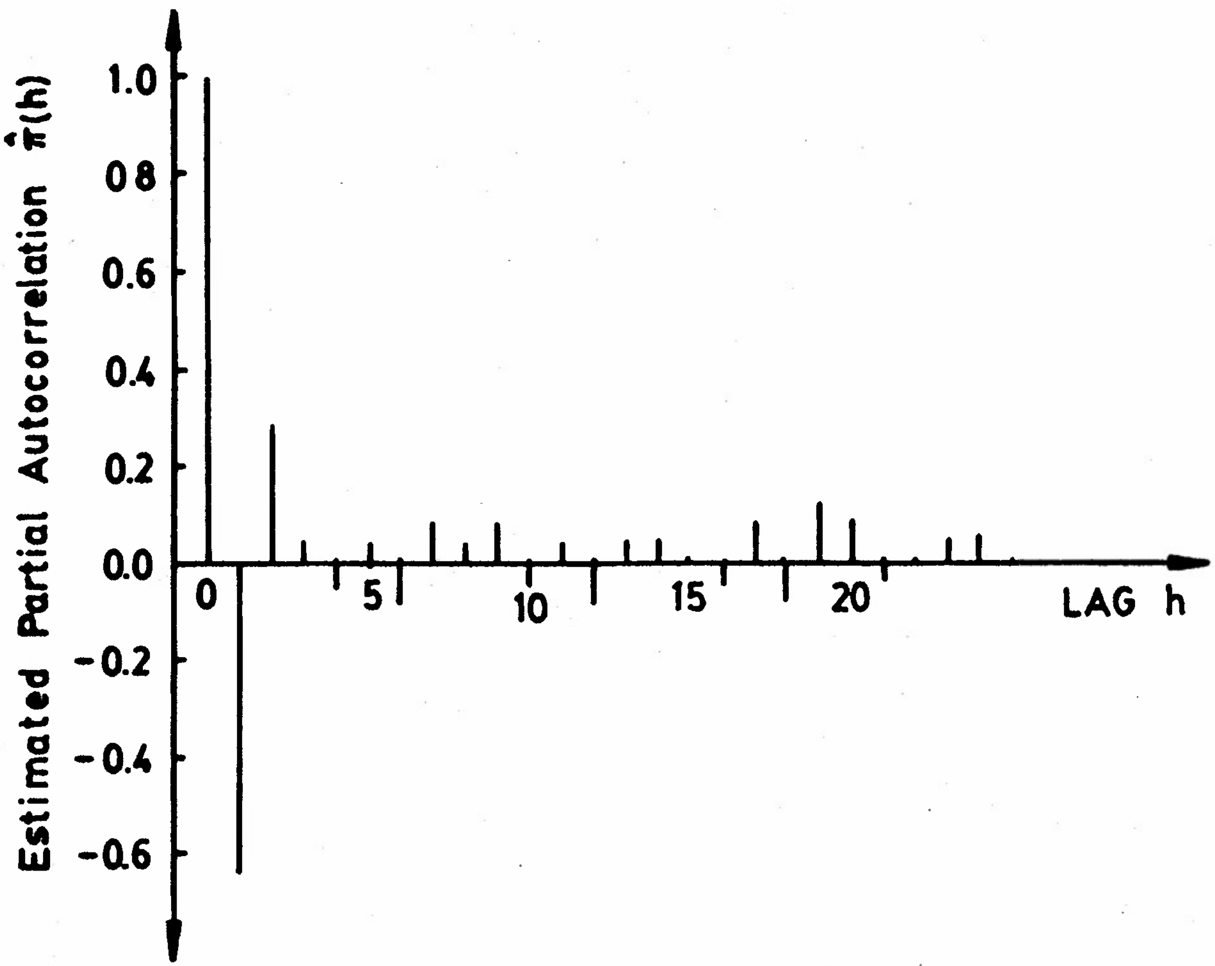

Figure 8.11. A plot of the estimated partial autocorrelation function $\hat{\pi}(h)$ versus h, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(2) process with $\beta_1 = .5$ and $\beta_2 = -.2$.

8.6 An Explanation of the Fluctuations in Autoregressive Processes

A typical time series described by an autoregressive process fluctuates up and down with oscillations which are not regular, but whose average length depends on the nature of the underlying difference equation. We can offer an explanation of these fluctuations by considering the representation

$$y_t = \sum_{r=0}^{\infty} \delta_r u_{t-r}$$

and noting that

$$y_{s+q} = \delta_0 u_{s+q} + \delta_1 u_{s+q-1} + \ldots + \delta_q u_s + \delta_{q+1} u_{s-1} + \ldots \quad .$$

Thus, a given u_s will influence a subsequent y_{s+q} via the coefficient δ_q. In Section 8.2 we have pointed out the circumstances under which the coefficients δ_q oscillate, causing fluctuations of the successive y_r's. To illustrate this point we show in Figure 8.12, a plot of a time series comprising of 100 observations generated on a computer by an AR(2) process

$$y_t - .9y_{t-1} + .4y_{t-2} = u_t \, , \qquad t = 3,4,\ldots, \quad .$$

This is the process considered in Sections 8.3.2, and 8.5.2. Since $\beta_1(=-.9)$ and $\beta_2(=.4)$ are such that $\beta_1^2 < 4\beta_2$, the roots of the associated polynomial equation are complex with $\theta = 44.68^\circ$, and $\alpha = .634$ (Section 8.2.1). Thus the coefficients δ_r will have a damped

sinusoidal behavior of the type indicated in Figure 8.2. Substituting the above values of θ and α in (8.18), we observe that $\delta_1 = .899$, $\delta_2 = .409$, $\delta_3 = .008$, $\delta_4 = -.156$, $\delta_5 = -.144$, $\delta_6 = -.06$, $\delta_7 = -.003$, $\delta_8 = .025$, $\delta_9 = .023$, and $\delta_{10} = .011$; the remaining values of δ_r, for $r \geq 1$ are all less than .002 and are thus essentially 0. It is because of the above behavior of the δ_r's that the observations y_t fluctuate up and down about 0 with an average length of oscillation of about length 10 - see Figure 8.12. It is also useful to keep in mind Figure 8.7, the estimated autocorrelation function of the generated series, and note that the estimated autocorrelations for lags greather than 10 are, barring sampling variability, essentially small.

8.7 Autoregressive Processes with Independent Variables

Suppose that there are m independent variables $z_{1t},\dots,z_{mt}$ that are known to affect the time series $\{y_t\}$, $t = 1,2,\dots,$ being investigated. The effect of these m variables can be incorporated into an AR(p) model, by writing

$$\text{(8.30)} \qquad \sum_{r=0}^{p} \beta_r y_{t-r} + \sum_{i=1}^{m} \gamma_i z_{it} = u_t \, , \qquad t = p+1, \dots \, ,$$

where $\gamma_1,\dots,\gamma_m$ are constants. It is of interest to compare the model (8.30) with the classicial regression model (2.1) of Part I.

Let $x_1,\dots,x_p$ be the roots of the associated polynomial equation of the stochastic difference equation $\sum_{r=0}^{p} \beta_r y_{t-r} = u_t$, $t = p+1,\dots,$.

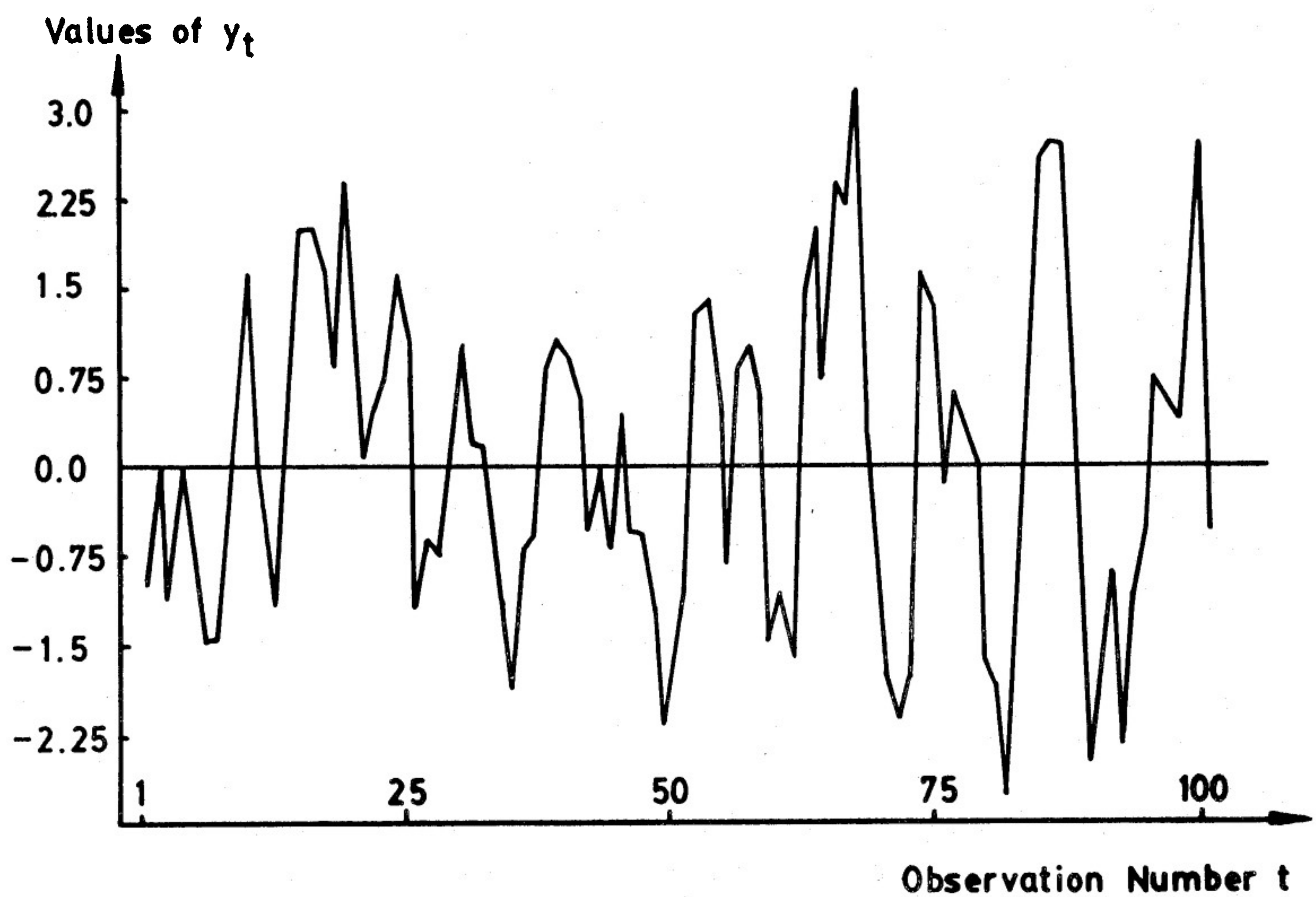

Figure 8.12. A plot showing the behavior of 100 computer generated observations from an AR(2) process $y_t - .9y_{t-1} + .4y_{t-2} = u_t$.

Then, using the forward lag operator $\wp$, where $\wp^s y_t = y_{t+s}$, $\sum_{r=0}^{p} \beta_r y_{t-r}$ can also be written as

$$\sum_{r=0}^{p} \beta_r y_{t-r} = \sum_{r=0}^{p} \beta_r \wp^{p-r} y_{t-p} = \prod_{i=1}^{p} (\wp - x_i) y_{t-p} \quad . \tag{8.31}$$

In terms of the operator $\mathcal{L}$, the above becomes

$$\left(\sum_{s=0}^{p} \beta_s \mathcal{L}^s\right) y_t = \prod_{i=1}^{p} (\wp - x_i) y_{t-p} \quad ,$$

or

$$\left(\sum_{s=0}^{p} \beta_s \mathcal{L}^s\right)^{-1} y_{t-p} = \left(\prod_{i=1}^{p} (\wp - x_i)\right)^{-1} y_t \quad .$$

Thus, for $t = p+1, p+2, \ldots,$ the model (8.30) becomes

$$\begin{aligned} u_t &= \prod_{j=1}^{p} (\wp - x_j) y_{t-p} + \sum_{i=1}^{m} \gamma_i z_{it} \\ &= \prod_{j=1}^{p} (\wp - x_j) [y_{t-p} + \sum_{i=1}^{m} \gamma_i \left(\prod_{j=1}^{p} (\wp - x_j)\right)^{-1} z_{it}] \\ &= \prod_{j=1}^{p} (\wp - x_j) [y_{t-p} + \sum_{i=1}^{m} \gamma_i \left(\sum_{s=0}^{p} \beta_s \mathcal{L}^s\right)^{-1} z_{i,t-p}] \quad . \end{aligned}$$

Before proceeding further, it is helpful to recall a result that we have encountered in Section 8.1.1, namely, that

$$\left(\sum_{r=0}^{p} \beta_r \mathcal{L}^r\right)^{-1} = \sum_{r=0}^{\infty} \delta_r \mathcal{L}^r \quad ,$$

where the δ_r's are the coefficients in the equality

$$(\sum_{r=0}^{p} \beta_r z^r)^{-1} = \sum_{r=0}^{\infty} \delta_r z^r \;;$$

see (8.8). Using the above, we now write u_t as

$$u_t = \prod_{j=1}^{p} (\wp - x_j)[y_{t-p} + \sum_{i=1}^{m} \gamma_i \sum_{s=0}^{\infty} \delta_s \mathcal{L}^s z_{i,t-p}]$$

$$= \prod_{j=1}^{p} (\wp - x_j)[y_{t-p} + \sum_{i=1}^{m} \delta_s z_{i,t-p-s}] \;,$$

or

(8.32) $$u_t = \sum_{r=0}^{p} \beta_r [y_{t-r} + \sum_{i=1}^{m} \gamma_i \sum_{s=0}^{\infty} \delta_s z_{i,t-r-s}] \;,$$

since

$$\prod_{j=1}^{p} (\wp - x_j) y_{t-p} = \sum_{r=0}^{p} \beta_r y_{t-r} \;.$$

A special case of the above is our AR(p) model (8.1). To see this, suppose that $m = 1$, and that $Z_{1t} = 1$, for all t. Then (8.32) becomes

(8.33) $$\sum_{r=0}^{p} \beta_r (y_{t-r} - \mu) = u_t \;, \qquad t = p+1,\ p+2, \ldots \;,$$

where $\mu = -\gamma_1 \sum_{s=0}^{\infty} \delta_s$.

8.8 Stationary Autoregressive Processes Whose Associated Polynomial Equations Have at Least One Root Equal to 1

Much of our discussion thus far, has been based on the requirement that all the roots of the associated polynomial equation $\sum_{r=0}^{p} \beta_r x^r = 0$ of an AR(p) process $\sum_{r=0}^{p} \beta_r \mathcal{L}^r y_t = u_t$, $t = p+1, p+2, \ldots,$ be less than 1 in absolute value. We have also assumed that the sequence of random variables $y_1, y_2, \ldots$ described by the AR(p) process be stationary. In this section, we investigate the implications of allowing the absolute value of one or more roots of the associated polynomial equation take a value equal to 1 and still maintain the requirement that the underlying sequence of random variables be stationary.

We begin by considering a stationary autoregressive process of order 1 with its single root taking a value 1; thus we have

$$y_t = y_{t-1} + u_t \,, \qquad t = 2,3,\ldots \,,$$

or

$$\Delta y_{t-1} = u_t \,, \qquad t = 2,3,\ldots \,.$$

Thus the first difference of our autoregressive process of order 1 with its single root equal to 1, is described by an innovation process. This latter process is stationary. We let $\mathcal{E}u_t = 0$, and $\mathcal{E}u_t^2 = \sigma^2$ for all values of t.

Now for all $s > 0$, we note that

$$y_t - y_{t-s} = u_t + u_{t-1} + \ldots + u_{t-s+1}$$

so that

$$\mathcal{E}(y_t-y_{t-s})^2 = \mathcal{E}y_t^2 + \mathcal{E}y_{t-s}^2 - 2\mathcal{E}y_t y_{t-s} = s\sigma^2 .$$

Since our sequence $\{y_t\}$ is stationary, $\mathcal{E}y_t^2 = \mathcal{E}y_{t-s}^2$, and so

$$\mathcal{E}y_t y_{t-s} = \sigma(s) = \mathcal{E}y_t^2 - \frac{s\sigma^2}{2} , \qquad s = 1,2,\ldots, .$$

The above result can hold for all $s > 0$ only if $\sigma^2 = 0$, in which case $y_t = y_{t-s}$, with probability 1.

To generalize, we consider a stationary autoregressive process of order p , $p > 1$, and allow one root, say x_1 to equal 1 , and require that the other $p-1$ roots are less than 1 in absolute value; that is, $|x_i| < 1$, $i = 2,\ldots,p$. Following (8.31), we may write our stationary AR(p) process as

$$(\wp-1)\prod_{i=2}^{p}(\wp - x_i)y_{t-p} = u_t .$$

If we let $\prod_{i=2}^{p}(\wp - x_i)y_{t-p} = z_{t-1}$, then our AR(p) process can be written as $(\wp-1)z_{t-1} = u_t$, or since $(\wp-1) = \Delta$, we have $\Delta z_{t-1} = u_t$.

It now follows from our previous discussion of the stationary AR(1) process with a single root equal to 1 that for all $s > 0$, $z_t = z_{t-s} = z$, say. Thus

$$\prod_{i=2}^{p}(\wp - x_i)y_{t-p} = z ,$$

and $y_t = \Sigma_{s=0}^{\infty} \delta_s z$, so that $y_t = y_{t-s}$, with probability 1 . We therefore have as

<u>Theorem 8.3</u>: If a stationary autoregressive process of order p has at least one root of its associated polynomial equation equal to 1 , then all values of the process are the same with probability 1 .

8.9 Some Linear Nonstationary Processes

We shall now introduce a type of nonstationary stochastic processes that are suitable for describing many empirical time series. Such series behave as though they have no fixed mean. The stochastic processes introduced here, are within the general structure of autoregressive processes.

Suppose that a *nonstationary* sequence of random variables y_1, y_2, ... is described by a stochastic difference equation of order p+d , so that $(\wp^{p+d} + \beta_1 \wp^{p+d-1} + \dots + \beta_{p+d} \wp^0) y_{t-p-d} = u_t$, $t = p+d+1$, p+d+2,..., or equivalently, the sequence is described by an autoregressive process of order p+d , where

$$\sum_{r=0}^{p+d} \beta_r \mathscr{L}^r y_t = u_t , \qquad t = p+d+1,\ p+d+2,\dots .$$

The associated polynomial equation $\sum_{r=0}^{p+d} \beta_r x^{p+d-r} = 0$ of the above process has p+d roots $x_1,\dots,x_p,x_{p+1},\dots,x_{p+d}$. Suppose that d of these roots, say $x_{p+1},\dots,x_{p+d}$ are exactly <u>equal</u> to 1 , and that the remaining $p(\geq 1)$ roots $x_1,\dots,x_p$, are <u>less</u> than 1 in absolute value. Then, following (8.31), we can write our AR(p+d) process as

$$\prod_{i=1}^{p} (\wp-x_i)(\wp-1)^d y_{t-p-d} = u_t$$

or

$$(8.34) \qquad \prod_{i=1}^{p} (\wp-x_i)\ \Delta^d y_{t-p-d} = u_t\ , \qquad t = p+d+1,\ p+d+2,\ldots\ ,$$

since $\wp-1 = \Delta$.

If we let $w_{t-p-d} = \Delta^d y_{t-p-d}$, and assume that the differenced sequence $\{w_{t-p-d}\}$, $t = p+d+1,\ p+d+2,\ldots,$ is <u>stationary</u>, then for $p \geq 1$ (8.34) becomes

$$(8.35) \qquad \prod_{i=1}^{p} (\wp-x_i) w_{t-p-d} = u_t\ , \qquad t = p+d+1,\ p+d+2,\ldots, \quad .$$

Thus for $p \geq 1$ our model for the nonstationary sequence $\{y_t\}$, $t = 1,2,\ldots,$ is one for which $\{w_{t-p-d}\}$, $t = p+d+1$, $p+d+2,\ldots,$ its d-th difference is described by a *stationary* autoregressive process of order p . Since the roots $x_1,\ldots,x_p$, are assumed to be less than 1 in absolute value, all our previous results for stationary autoregressive processes are also applicable for the model (8.35).

When $p = 0$, and $d = 1$, the first difference of our nonstationary sequence $\{y_t\}$ is described by an innovation process $\{u_t\}$ which by definition is always stationary; see Section 8.8.

Note that if the original series $\{y_t\}$, $t = 1,2,\ldots,$ consists of n observations, then the differenced series $\{w_t\}$ will consist of $n-d$ observations. Since $w_{t-p-d} = \Delta^d y_{t-p-d}$, $t = p+d+1,\ p+d+2,\ldots,$ we write $y_{t-p-d} = \Delta^{-d} w_{t-p-d}$, where the notation Δ^{-d} needs to be explained. For this purpose we set $d = 1$, substitute the values $t = p+2,\ p+3,\ldots,$ in the telescoped series $w_{t-p-1} = y_{t-p} - y_{t-p-1}$, and

observe that we can write $y_{t-p} = w_{t-p-1} + w_{t-p-2} + \ldots + w_1 + y_1$, for any $t \geq p+2$. The operator Δ^{-1} therefore represents summation or integration - the reverse of differencing - and it is for this reason that we say that the sequence $\{y_t\}$, $t = 1,2,\ldots,$ is described as an integrated autoregressive process of order p. An explanation for Δ^{-d}, $d \geq 2$, follows by an analogous argument.

8.9.1 Behavior of Estimated Covariance Functions of Integrated Autoregressive Processes and Processes with an Underlying Trend

Since the associated polynomial equation of the process described by the difference equation

$$(8.36) \qquad (\wp^{p+d} + \beta_1 \wp^{p+d-1} + \ldots + \beta_{p+d}\wp^0)y_{t-p-d} = u_t\,, \qquad t = p+d+1,\ldots\,,$$

is $\sum_{r=0}^{p+d} \beta_r x^{p+d-r} = 0$, it follows from (8.22) that if the roots $x_1, \ldots, x_{p+d}$ are distinct, $|x_j| < 1$, $j = 1,\ldots,p+d$, and $\beta_{p+d} \neq 0$, $\sigma(h)$, the covariance between y_t and $y_{t\pm h}$, is

$$\sigma(h) = \sum_{i=1}^{p+d} c_i x_i^h\,, \qquad h = 1-p-d,\ 2-p-d,\ldots,0\,,$$

where $c_1,\ldots,c_{p+d}$ are coefficients.

Since the roots are assumed to be real, distinct, and less than 1 in absolute value, each x_i^h damps exponentially. If a pair of roots, say x_j and x_k are conjugate complex; then $\sigma(h)$ is a mixture of

damped exponentials and damped sine waves. Now suppose that one of the roots, say x_ℓ is close to 1, so that for some small number $\delta > 0$,

$$x_\ell = 1 - \delta .$$

Then using a first order Taylor's series expansion for $(1-\delta)^h$, we see that $(1-\delta)^h \approx 1 - h\delta$, for an arbitrary h, and so $c_\ell x_\ell^h$ contributes a term approximately $c_\ell(1-h\delta)$ to $\sigma(h)$. The term $c_\ell(1-h\delta)$ decreases linearly and slowly in h, for h not too large. Thus, the tendency of the estimated autocorrelation function to decrease linearly and slowly in h, indicates the possible presence of a root close to 1 (in absolute value), in (8.36), the associated polynomial equation of the process. When such is the case and the underlying series cannot be assumed stationary (for otherwise the result of Theorem 8.3 would come into effect), we may want to consider appropriate differences of the series and attempt to model these as a stationary autoregressive process; see the discussion following (8.35).

To illustrate the above issues, we generate on a computer 250 observations from an AR(1) process

$$y_t - .99y_{t-1} = u_t, \qquad t = 2,3,\ldots .$$

A time series plot of these 250 observations is shown in Figure 8.13. An examination of this plot reveals that the series behaves as though it has no fixed mean. This is to be expected since $\beta_1 = -.99$ being close to 1 in absolute value makes the nonstationarity of the generated series a likely possibility. In Table 8.7 we give values of $r(h)$,

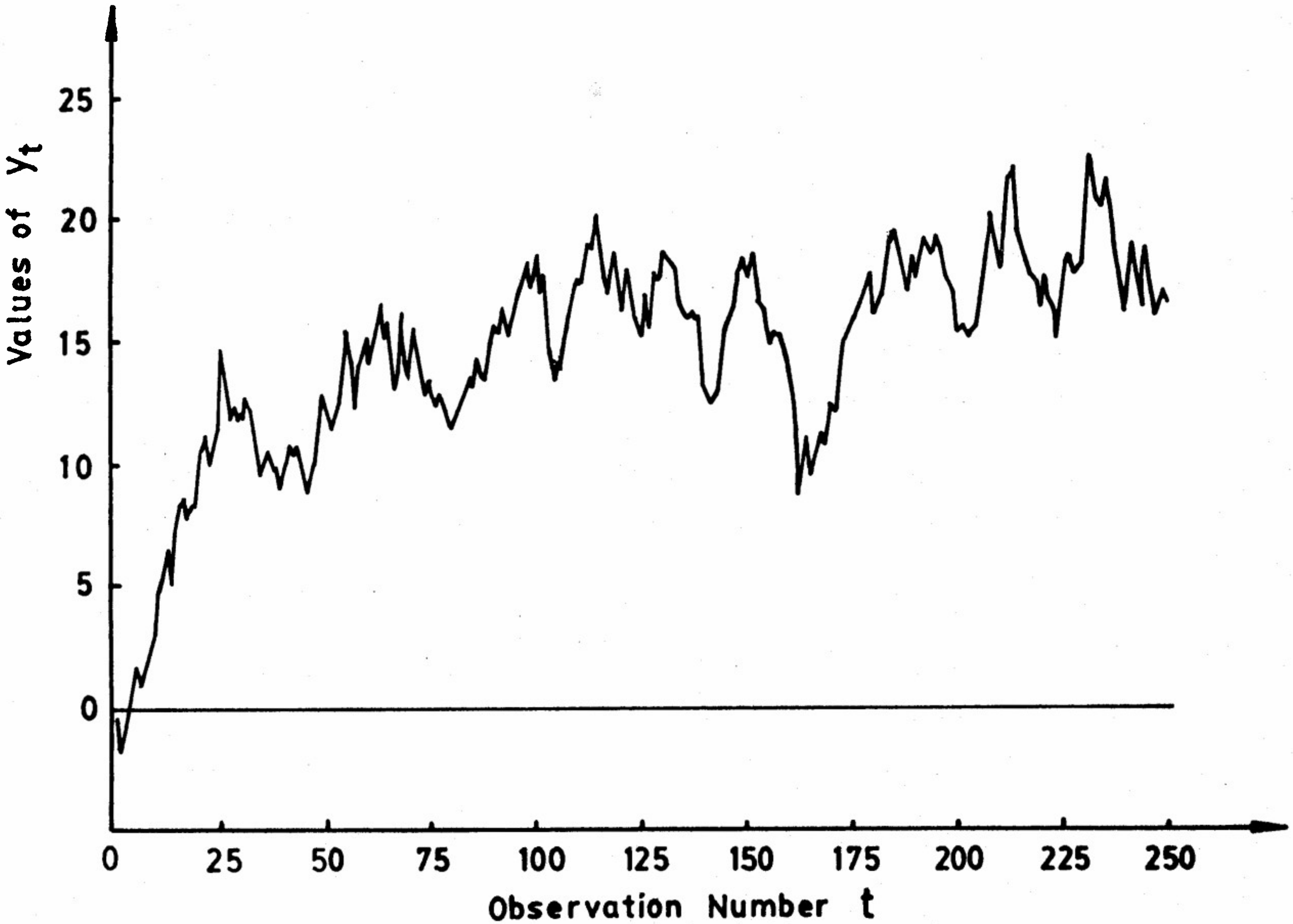

Figure 8.13. A plot showing the behavior of 250 computer generated observations from an AR(1) process $y_t - .99y_{t-1} = u_t$.

Table 8.7

Values of the estimated autocorrelation function $r(h)$, $h = 0,1,\ldots,50$, based on 250 computer generated observations from an AR(1) process with $\beta_1 = -.99$.

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	
Value of r(h)	1	.95	.89	.84	.79	.75	.70	.64	.59	.55	.51	.48	
Lag h	12	13	14	15	16	17	18	19	20	21	22	23	24
Value of r(h)	.45	.43	.40	.39	.38	.36	.35	.34	.33	.32	.32	.31	.30
Lag h	25	26	27	28	29	30	31	32	33	34	35	36	37
Value of r(h)	.28	.29	.30	.30	.30	.30	.30	.30	.31	.31	.31	.30	.29
Lag h	38	39	40	41	42	43	44	45	46	47	48	49	50
Value of r(h)	.27	.25	.23	.20	.19	.17	.15	.13	.12	.09	.07	.06	.06

the estimated autocorrelation function, for $h = 0,1,\ldots,50$, and in Figure 8.14 we show a plot of $r(h)$ versus h. As is to be expected, this plot conspicuously shows the slow, and almost linear decay, of the estimated autocorrelation function.

Since the estimated autocorrelation function of the generated series decays slowly and linearly, we consider $w_t = y_{t+1} - y_t$, $t = 1,2,\ldots$, the first difference of the generated series, and investigate the behavior of its estimated autocorrelation function. Recall, that if β_1 were to be exactly equal to -1, then the w_t's would be described by an innovation process whose autocorrelations at all lags other than 0, is zero.

In Figure 8.15, we show a plot of the time series generated by the w_t's, for $t = 1,2,\ldots,249$. In contrast to Figure 8.13, we see that the differenced series $\{w_t\}$ reveals fluctuations around a fixed mean of zero. In Table 8.8 we give values of $r(h)$, the estimated autocorrelation function of the w_t series, for $h = 0,1,\ldots,25$, and in Figure 8.16 we show a plot of $r(h)$ versus h. We contrast this plot in Figure 8.16 with that of Figure 8.14, and note that in the former, as is to be expected, the autocorrelations at lags other than 0 are, barring sampling variability, effectively zero.

8.9.2 The Covariance Function of Some Processes with an Underlying Trend

As a note of caution, it is not true that the tendency of the estimated autocorrelation function to decrease slowly necessarily implies that a root close to 1 exists. Such a tendency can also be

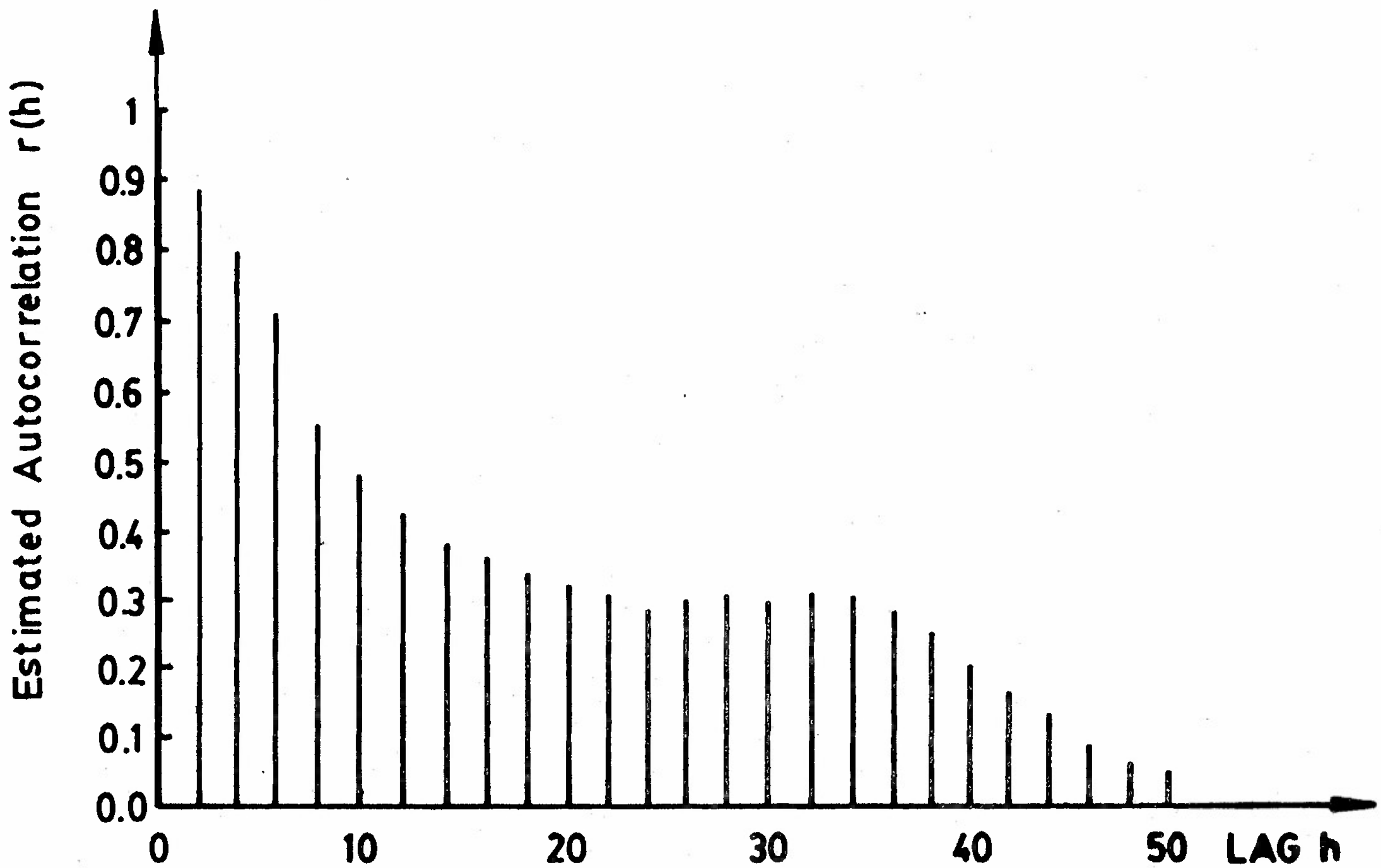

Figure 8.14. A plot of the estimated correlation function $r(h)$ versus h, $h = 0,1,\ldots,50$ based on 250 computer generated observations from an AR(1) process with $\beta_1 = -.99$.

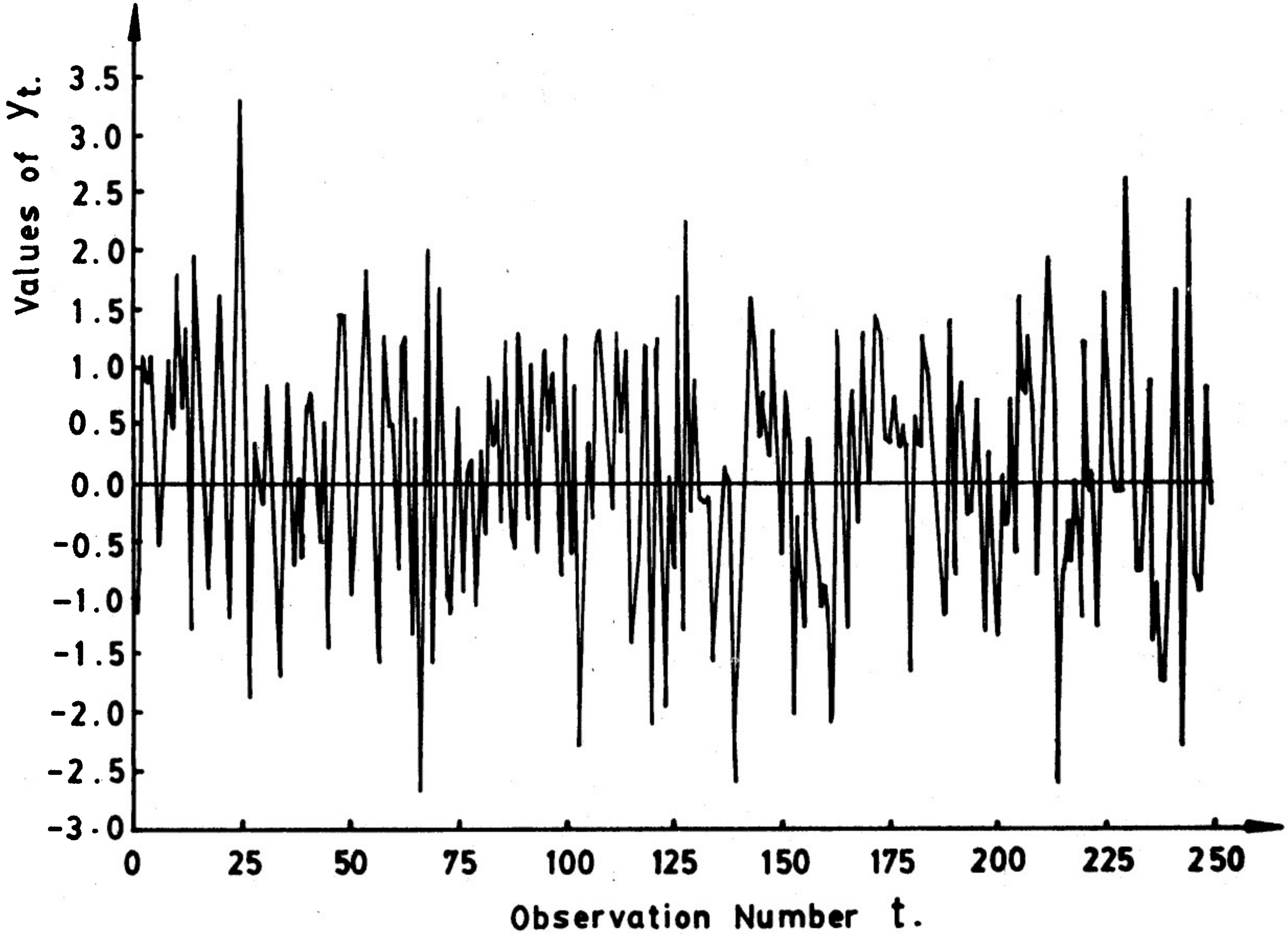

Figure 8.15. A plot showing the behavior of the first differences of the 250 computer generated observations from an AR(1) process with $\beta_1 = -.99$.

Table 8.8

Values of the estimated autocorrelation function $r(h)$, $h = 0,1,\dots,25$, for the first differences of the 250 computer generated observations from an AR(1) process with $\beta_1 = -.99$

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	1	.02	-.09	-.06	.04	.10	.03	-.01	-.14	-.06	.03	-.09	-.03
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of r(h)	-.05	-.01	-.03	.01	-.02	-.03	.07	-.01	-.01	.06	.03	.06	-.17

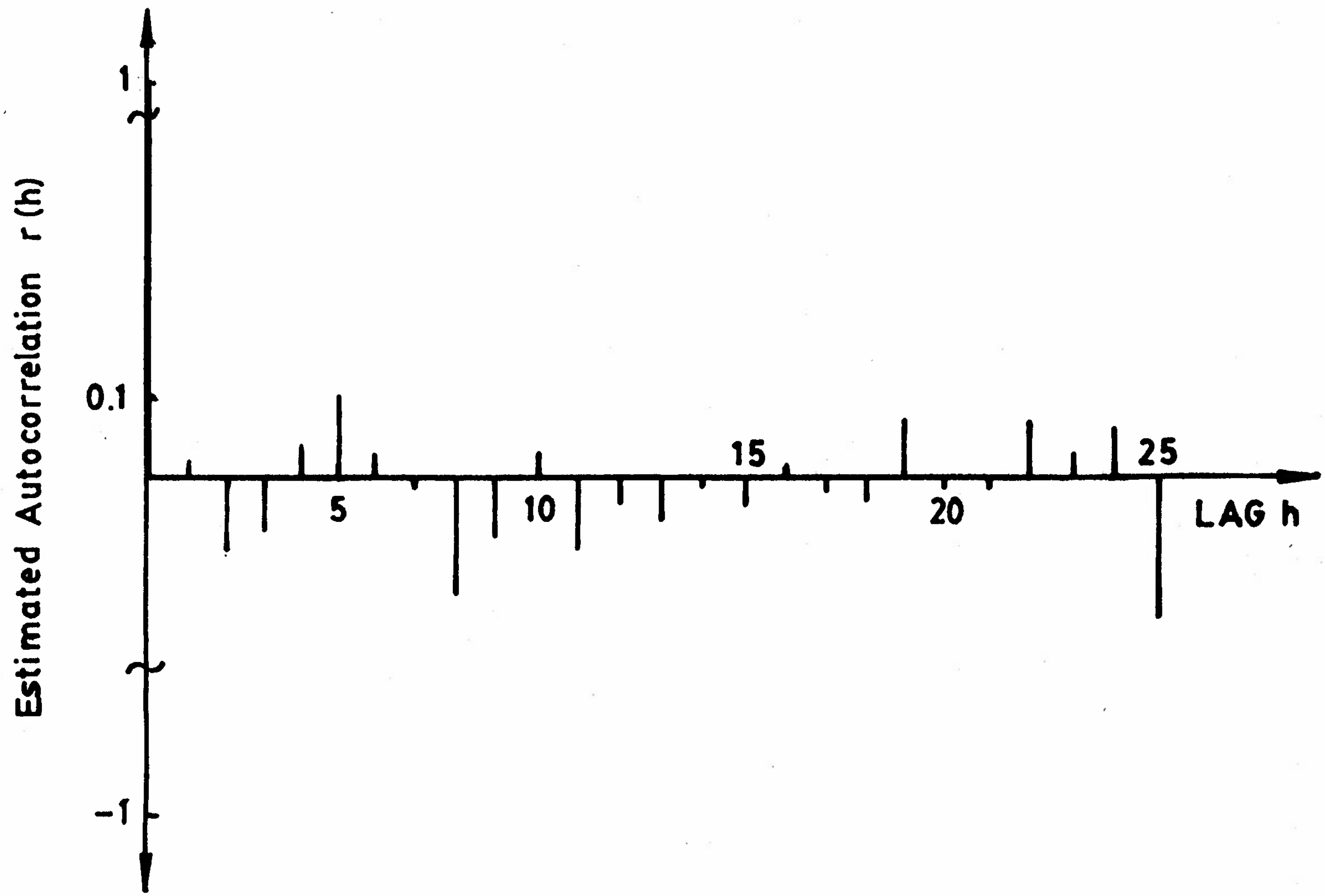

Figure 8.16. A plot of the estimated correlation function $r(h)$ versus h, $h = 0,1,\ldots,25$ of the first differences of the 250 computer generated observations from an AR(1) process with $\beta_1 = -.99$.

observed whenever there is an underlying trend in the series. To see why this is so, let us consider, for example, a process with an underlying linear trend of the form

$$y_t = u_t + t , \qquad t = 1,2,\ldots ,$$

where, as before, $u_1, u_2, \ldots$ are independent and identically distributed with mean 0 and variance σ^2.

Since $\mathcal{E} y_t = t$ for all values of t, it is easy to verify that for a series of length T the expected value of the sample mean of $y_1, \ldots, y_T$ is

$$\mathcal{E} \frac{1}{T} \sum_{t=1}^{T} y_t = \frac{T+1}{2} .$$

If we are to use the numerator of (7.4) to compute the covariance of the T observations at lag $h > 0$, then the theoretical quantity that is being estimated is

$$\mathcal{E} \frac{1}{T} \sum_{t=1}^{T-h} \left(y_t - \frac{T+1}{2}\right)\left(y_{t+h} - \frac{T+1}{2}\right) . \tag{8.37}$$

Using the fact that $\mathcal{E} y_t y_{t+h} = t(t+h)$, $h = 1,2,\ldots,$ and that $\sum_{t-1}^{T-h} t^2 = (T-h)(T-h+1)(2T-2h+1)/6$, we can show that for large values of T, (8.37) can be approximated by $(T-h)((T-h)^2 - 3h^2)/24$, which for small values of h decreases slowly in h.

An analogous conclusion can be drawn for other types of processes and other types of trends. Thus in practice, to investigate the nature of the series, by plotting it to see whether it exhibits an

underlying trend. If the underlying trend appears to be a polynomial, then, as discussed in Section 3.3, such a trend can be eliminated by taking an appropriate number of differences of the series.

We have thus seen that the differencing of an observed series may be motivated by two distinct considerations. The first enables us to model certain types of nonstationary sequence of random variables, via the mechanism of a integrated autoregressive processes, whereas the second enables us to eliminate the presence of an underlying polynomial trend in a series.

To illustrate the effects of a linear trend on the estimated autocorrelation function of a series, we generate on the computer, 250 observations from an autoregressive process of order 1 with $\beta_1 = -.5$, and for which a linear trend term is added. Note that this is the same series considered in Sections 8.3.2 and 8.5.2, except that the inclusion of a linear trend term makes the generated series nonstationary.

In Table 8.9 we give $r(h)$, the values of the estimated autocorrelation function, for $h = 0,1,\ldots,25$, for the 250 computer generated observations described above. In Figure 8.17, we plot $r(h)$ versus h. This plot clearly shows the very slow decay of the estimated autocorrelation function. The estimated autocorrelation and partial autocorrelation functions of the first difference of this series will reveal a behavior analogous to those of Figures 8.6 and 8.9, since by differencing the series we would have eliminated the linear trend.

An example of some real life data with a trend, and for which the estimated autocorrelation function decreases linearly and slowly is given in the next section.

Table 8.9

Values of the estimated autocorrelation function $r(h)$, $h = 0,1,\ldots,25$, based on 250 computer generated observations from an AR(1) process with $\beta_1 = -.5$ and a linear trend term added to it

Lag h	0	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	1	.82	.72	.65	.61	.58	.57	.58	.58	.57	.54	.52	.50
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of r(h)	.53	.50	.49	.48	.47	.46	.48	.47	.46	.46	.47	.47	.45

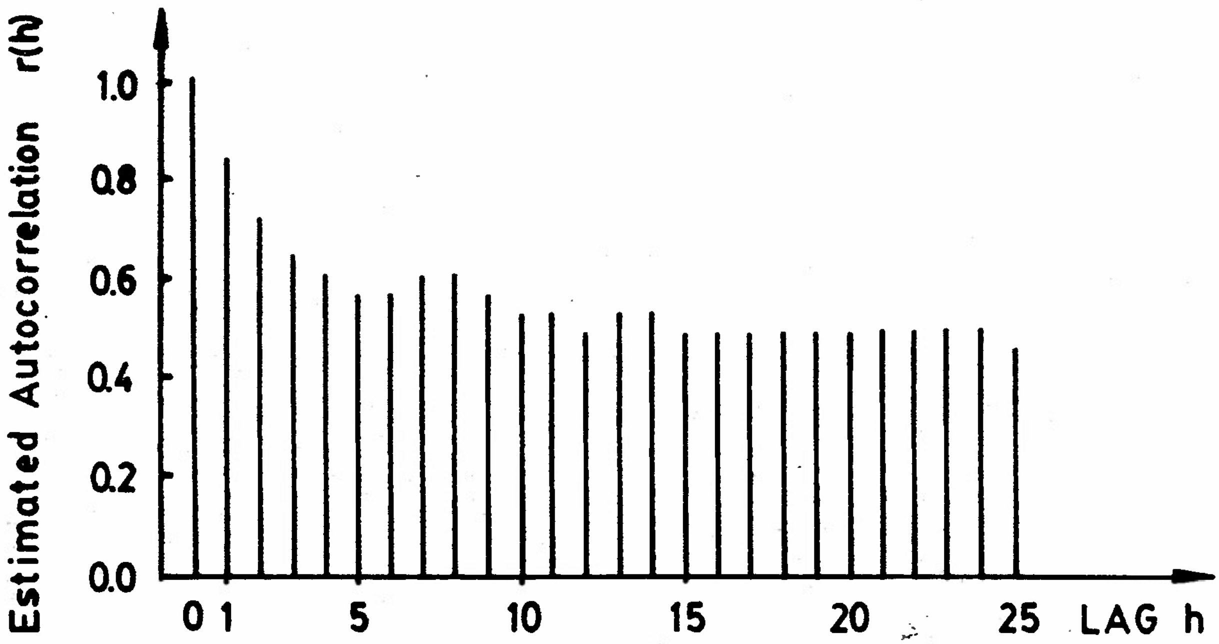

Figure 8.17. A plot of the estimated autocorrelation function $r(h)$ versus h, $h=0,1,\ldots,25$, based on 250 computer generated observations from an AR(1) process with $\beta_1=-.5$, and a linear trend term added to it.

8.9.3 Behavior of the Estimated Autocorrelation Function of a Real Life Nonstationary Time Series

In Figure 8.18, we show a plot of G_t, the US gross national product (GNP) in billions of US dollars, for the years $t = 1920$ through $t = 1979$ (Source: Bureau of Economic Analysis, U.S. Department of Commerce, Washington, D.C.). This plot indicates that the GNP series is a nonstationary one, it being increasing (approximately) exponentially over time. Thus it appears reasonable to first take the natural logarithms of the GNP, $\ln G_t$. A plot of $y_t = \ln G_t \times 1000$ versus t is shown in Figure 8.19; the actual values of y_t are given in Table 8.10. Figure 8.19 shows that the y_t series is also not stationary, it being increasing (approximately) linearly in t. The estimated autocorrelation function of the y_t series, for lags 0 through 20, is shown in Figure 8.20. Because this estimated autocorrelation function decreases linearly and slowly, we consider the first differences of the y_t series, $w_t = \Delta y_t = y_{t+1} - y_t$, $t = 1,2,\ldots,59$. A plot of w_t versus t is shown in Figure 8.21; the actual values of w_t are also given in Table 8.10. From Figure 8.21 we see that whereas the w_t series appears to have a constant level (mean), its fluctuations in the earlier years, 1920-1947, appear to be more erratic than the fluctuations in the latter years, 1947-1979. A possible explanation for this behavior is that "automatic stabilizers" such as unemployment insurance, workmens compensation, etc., which were introduced into the economy as of 1947, tend to make the GNP less erratic. Plots of the estimated autocorrelation and partial autocorrelation functions of the w_t series are shown in Figures 8.22 and

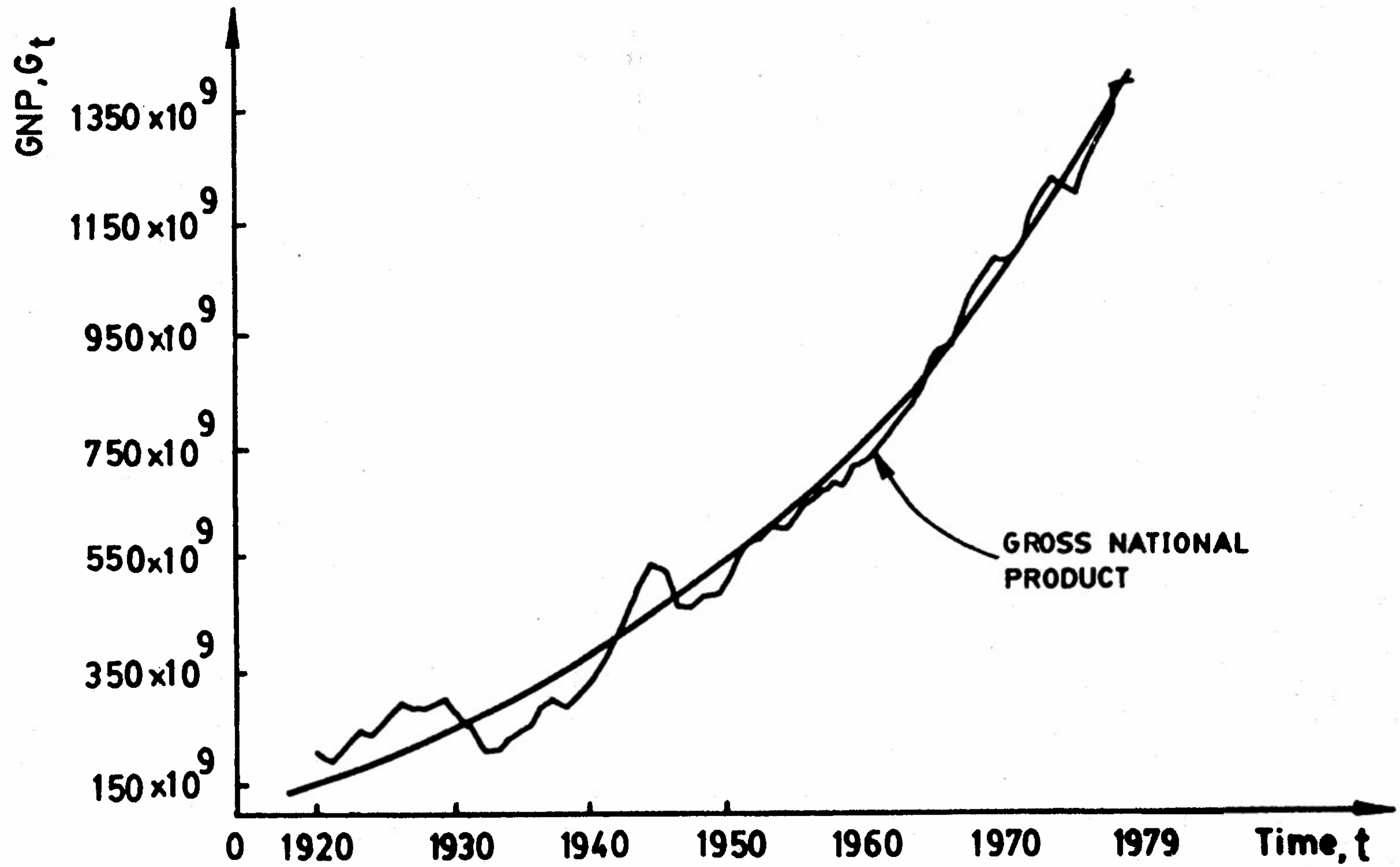

Figure 8.18. A plot of G_t the US Gross National Product in billions of US Dollars, for 1920 through 1979.

(The quadratic curve superimposed on the Gross National Product indicates the nature of the trend.)

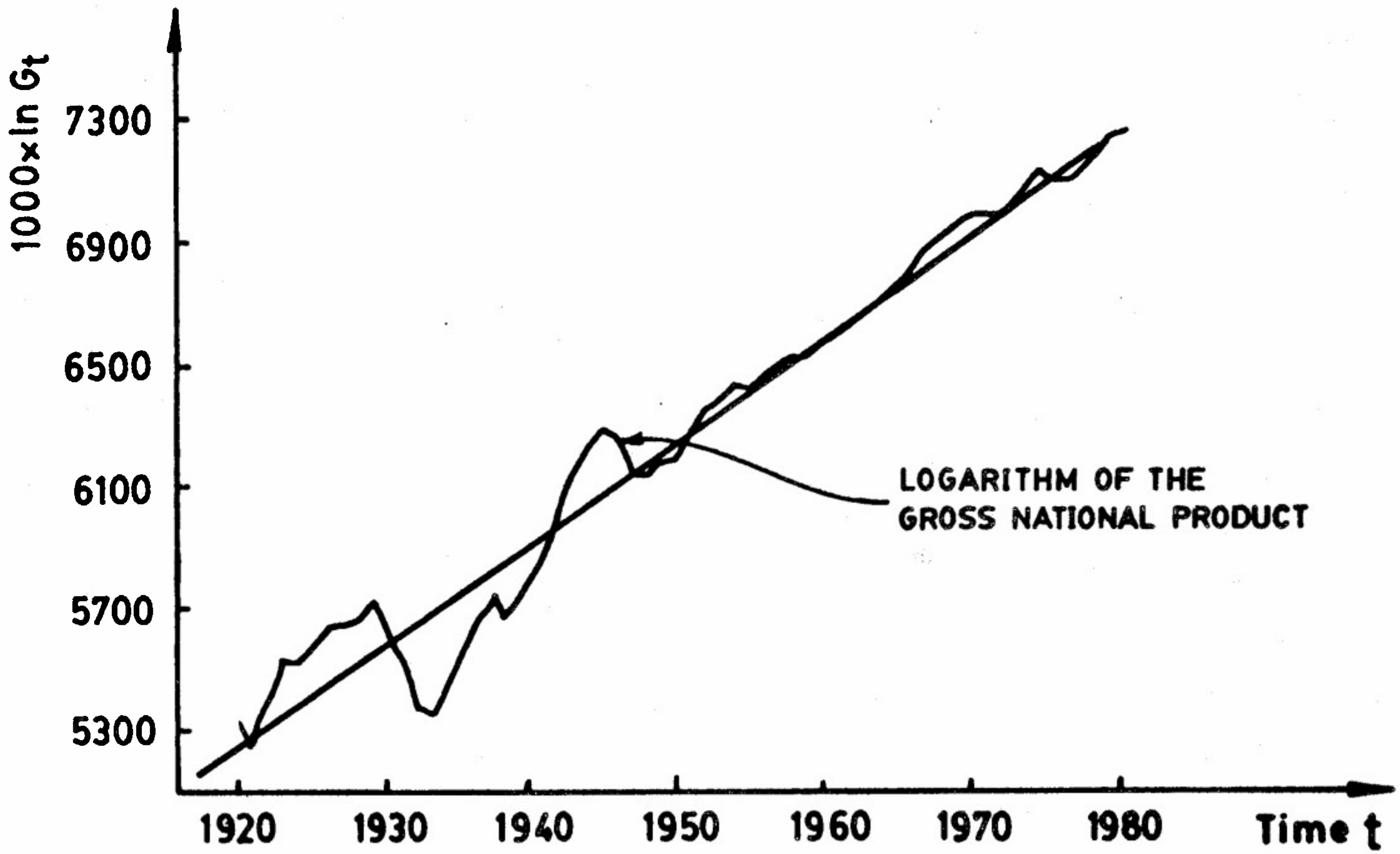

Figure 8.19. A plot of the natural logarithm of the Gross National Product in billions of US Dollars times 1000 for 1920 through 1979.

(The straight line superimposed on the logarithm of the Gross National Product indicates the nature of the trend.)

Table 8.10

Values of y_t, the natural logarithm of the Gross National Product × 1000, for $t = 1920$ through 1979 and $w_t = y_{t+1} - y_t$, the first forward difference of y_t.

Year	$y_t = 1000 \times \ln G_t$	First Difference $w_t = y_{t+1} - y_t$	Year	$y_t = 1000 \times \ln G_t$	First Difference $w_t = y_{t+1} - y_t$
1920	5346.63	-91.22	1950	6279.46	77.52
1921	5255.41	146.82	1951	6356.98	37.45
1922	5402.23	114.02	1952	6394.43	38.19
1923	5516.25	-2.42	1953	6432.62	-13.11
1924	5513.83	80.88	1954	6419.51	64.82
1925	5594.71	57.43	1955	6484.33	21.16
1926	5652.14	-1.06	1956	6505.49	17.93
1927	5651.08	5.61	1957	6523.42	-2.06
1928	5656.69	64.60	1958	6521.36	58.45
1929	5751.29	-104.16	1959	6579.81	22.51
1930	5617.13	-80.58	1960	6602.32	24.80
1931	5536.55	-160.35	1961	6627.12	56.37
1932	5376.20	-18.67	1962	6683.49	38.78
1933	5357.53	86.18	1963	6722.27	51.27

(continued on page 90)

Table 8.10 (continued)

Year	$y_t = 1000 \times \ln G_t$	First Difference $w_t = y_{t+1} - y_t$	Year	$y_t = 1000 \times \ln G_t$	First Difference $w_t = y_{t+1} - y_t$
1934	5543.72	94.01	1964	6773.54	57.23
1935	5537.73	130.08	1965	6830.77	57.80
1936	5667.81	51.51	1966	6888.57	26.85
1937	5719.33	-52.21	1967	6915.43	42.83
1938	5667.12	82.27	1968	6958.26	25.34
1939	5749.39	81.32	1969	6983.60	-3.25
1940	5830.71	149.19	1970	6980.35	29.51
1941	5979.90	121.54	1971	7009.86	55.84
1942	6101.44	123.91	1972	7065.70	51.83
1943	6225.35	69.36	1973	7117.53	-18.58
1944	6294.71	-16.94	1974	7098.95	-6.96
1945	6277.77	-127.81	1975	7091.99	55.56
1946	6149.96	-0.85	1976	7147.56	47.40
1947	6149.11	40.59	1977	7194.96	48.70
1948	6189.70	6.13	1978	7243.66	22.54
1949	6195.83	83.63	1979	7266.20	

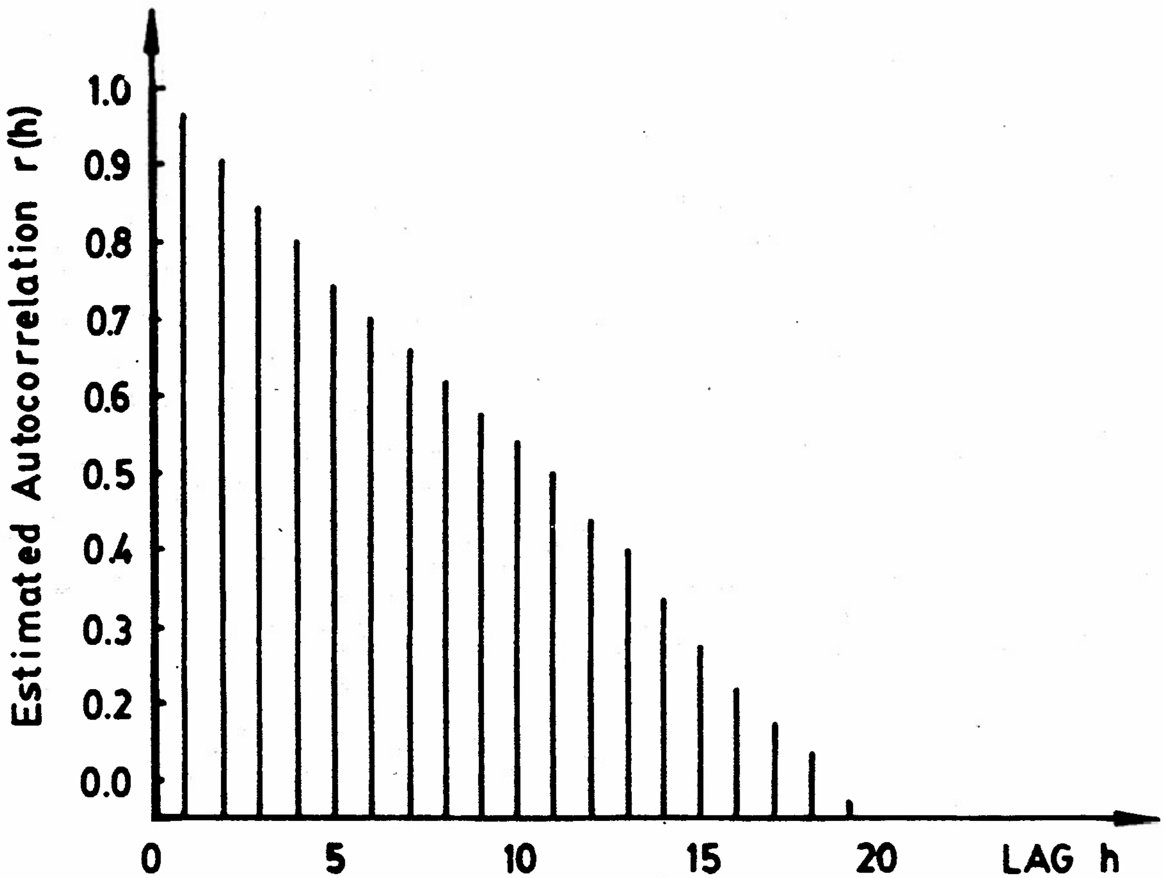

Figure 8.20. The estimated autocorrelation function of y_t, the Logarithm of GNP times 1000.

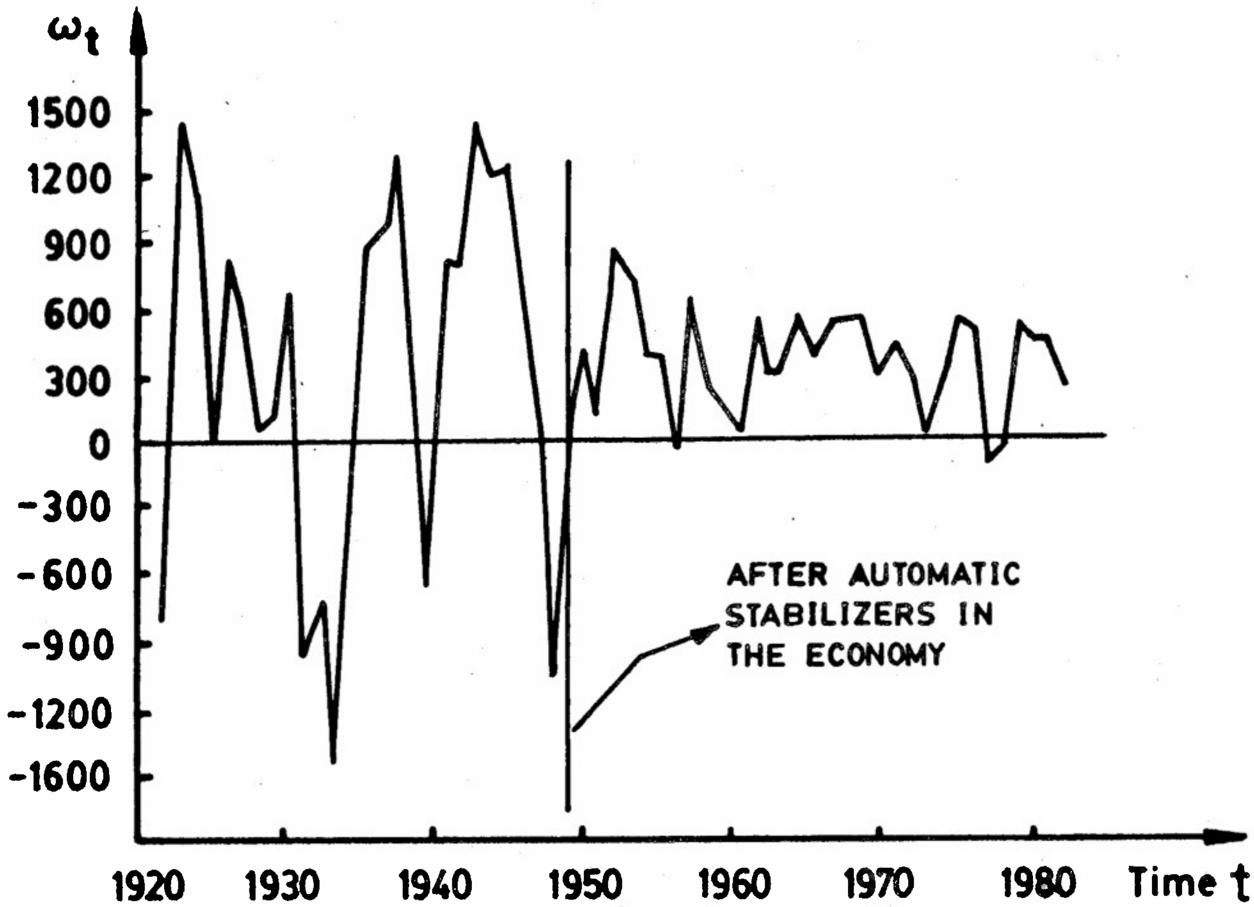

Figure 8.21. A plot of w_t, the first difference of $\ln GNP \times 1000$ versus time, for 1920 through 1979.

8.23, respectively. In contrast to Figure 8.19, we note from Figure 8.22 that the autocorrelation function at lag 1 is significantly different from 0 and takes smaller values for the other lags. However, it is difficult for us to claim in Figure 8.22, a pattern of either an exponential or a sinusoidal decay. A similar type difficulty is apparent in Figure 8.23. Thus it appears that the w_t series cannot be reasonably well described by an autoregressive process of the type discussed here; more complicated models to be presented later on, may be necessary for analyzing this data. Our main goal here, is to show Figure 8.20.

8.10 Forecasting (Prediction) for Stationary Autoregressive Processes

Suppose that a sequence of random variables $\{y_t\}$, $t=1,2,\ldots,$ can be described by an autoregressive process of order p

$$y_t + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \cdots + \beta_p y_{t-p} = u_t\,, \quad t = p+1,\ p+2, \ldots, \ .$$

Let $y^*_{t-1}, y^*_{t-2}, \ldots,$ be the observed values of the random variables $y_{t-1}, y_{t-2}, \ldots$. Our goal in this section is to determine a procedure which would give us a best, in the sense of minimum mean square error, forecast (predictor) of the unobserved variable y_t. We shall soon see that if the roots of the associated polynomial equation $\sum_{r=0}^{p} \beta_r x^{p-r} = 0$ of the AR(p) process given above are all less than 1 in absolute value, then the best predictor of y_t is indeed

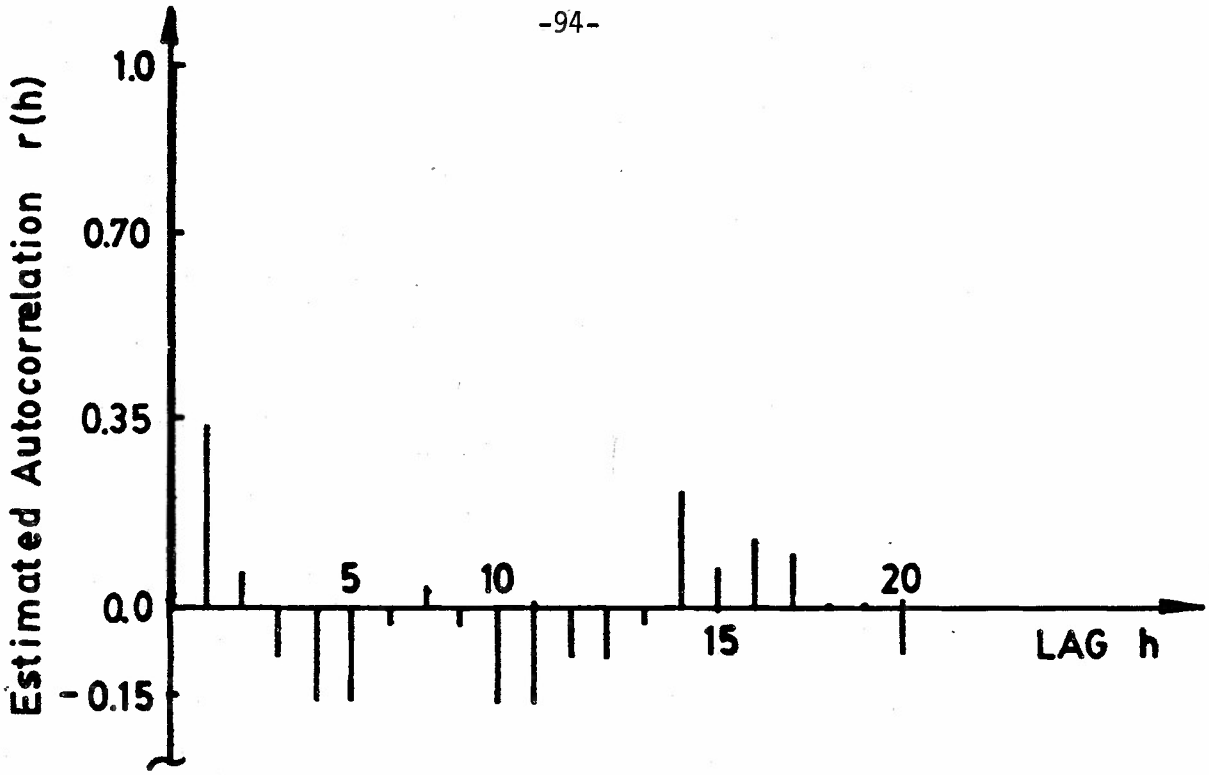

Figure 8.22. The estimated autocorrelation function of w_t, the first difference of lnGNP × 1000.

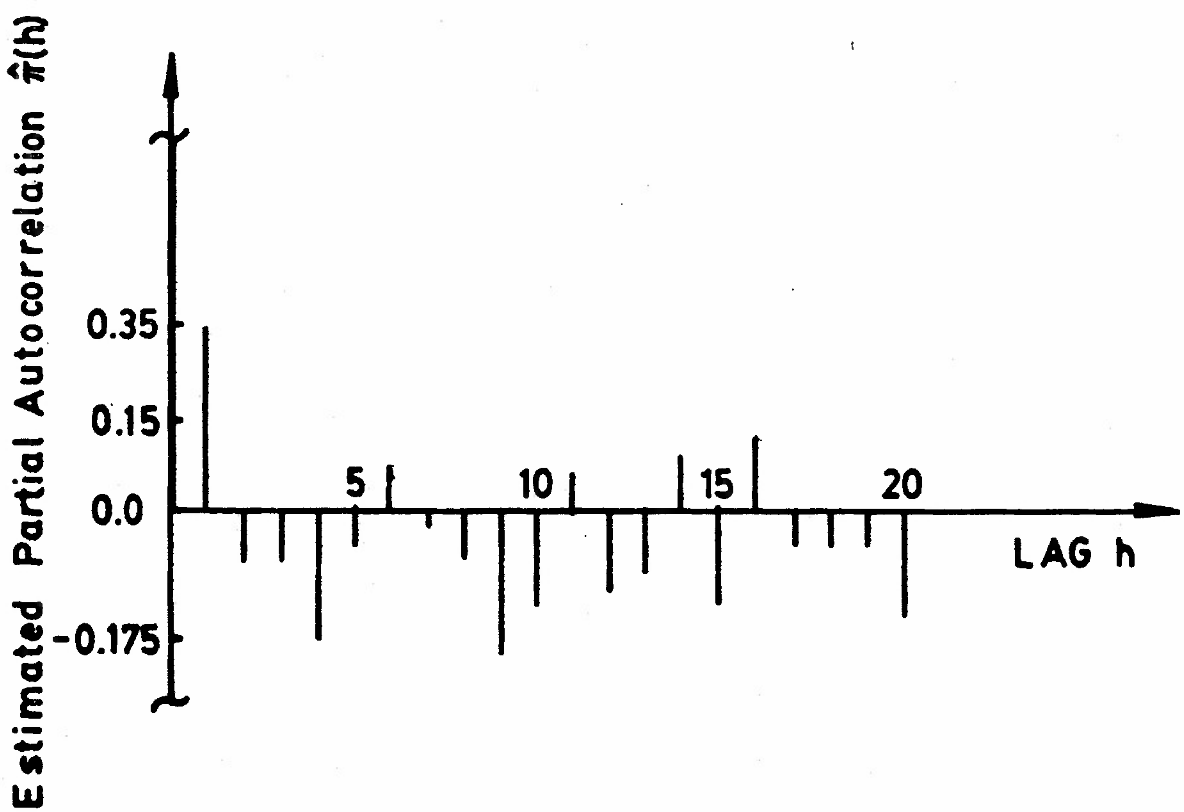

Figure 8.23. The estimated partial autocorrelation function of w_t, the first difference of lnGNP × 1000.

the natural quantity

$$-\beta_1 y^*_{t-1} - \beta_2 y^*_{t-2} - \cdots - \beta_p y^*_{t-p} \quad .$$

To see that the above is true, we first note that under the conditions of Theorem 8.1 in the equation

$$y_t = -\beta_1 y_{t-1} - \beta_2 y_{t-2} - \cdots - \beta_p y_{t-p} + u_t$$

u_t is independent of $y_{t-1}, y_{t-2}, \ldots$. Thus, the conditional expectation of y_t given that $y_{t-1} = y^*_{t-1}$, $y_{t-2} = y^*_{t-2}$, ... is

$$\mathcal{E}(y_t | y_{t-1} = y^*_{t-1}, y_{t-2} = y^*_{t-2}, \ldots,) = -\beta_1 y^*_{t-1} - \beta_2 y^*_{t-2} - \cdots - \beta_p y^*_{t-p} \ ;$$

the right-hand side of the above equation can be used to forecast y_t, given $y^*_{t-1}, y^*_{t-2}, \ldots,$.

Now let $f(y^*_{t-1}, y^*_{t-2}, \ldots,)$ be any other function of the previous values $y^*_{t-1}, y^*_{t-2}, \ldots,$ and suppose that we use $f(y^*_{t-1}, y^*_{t-2}, \ldots,)$ as a another forecast of y_t. Then the mean square error of $f(y^*_{t-1}, y^*_{t-2}, \ldots)$ as a predictor of y_t is

$$\begin{aligned}\mathcal{E}[f(y^*_{t-1}, \ldots) - y_t]^2 &= \mathcal{E}[f(y^*_{t-1}, \ldots) + \beta_1 y^*_{t-1} + \cdots + \beta_p y^*_{t-p} - u_t]^2 \\ &= \mathcal{E}u_t^2 + \mathcal{E}[f(y^*_{t-1}, \ldots) + \beta_1 y^*_{t-1} + \cdots + \beta_p y^*_{t-p}]^2 \ ,\end{aligned}$$

since u_t is independent of $y_{t-1}, y_{t-2}, \ldots$.

The above is minimized when $f(y^*_{t-1}, y^*_{t-2}, \ldots) = -\beta_1 y^*_{t-1} - \beta_2 y^*_{t-2} - \cdots - \beta_p y^*_{t-p}$.

In general, when the conditions of Theorem 8.1 are satisifed, the best (minimum mean square error) forecast of y_t given y_{t-s-1}, $y_{t-s-2},\ldots,$ $(s\geq 0)$ is $\mathcal{E}(y_t|y_{t-s-1}, y_{t-s-2},\ldots)$, the conditional expectation of y_t given $y_{t-s-1}, y_{t-s-2},\ldots,$ where $\mathcal{E}\{y_t|y_{t-s-1}, y_{t-s-2},\ldots\} = \alpha_{s1}y_{t-s-1} + \alpha_{s2}y_{t-s-2} + \ldots + \alpha_{sp}y_{t-s-p}$, and $\alpha_{s1} = \alpha^*_{s1},\ldots,\alpha_{sp} = \alpha^*_{sp}$ are given in equation (8.6).

8.11 Examples of Some Real Life Time Series Described by Autoregressive Processes

In this section our aim is to demonstrate the methods of the previous sections by considering some real life data which can be reasonably well described by autoregressive processes. It is entirely possible that the data can also be described by some of the other models introduced later on. However, at this point in time, it is convenient to introduce the data and use these to indicate the practical usefulness of autoregressive processes of a simple order.

8.11.1 The Weekly Rotary Rigs in Use Data*

A rotary rig is composed of five major components and costs upward of $500,000 each. It is used for drilling for oil and gas. The number of rotary rigs in use per week, by state, is an important component used in econometric models of the oil and gas industry. Such econometric models are of interest to the U.S. Department of Energy.

*This data and its description was given to us by Mrs. B. Volpe, Energy Information Office, U.S. Department of Energy, Washington, D.C. It has been abstracted from a Hughes Tool Company, Houston, Texas, report, entitled "Average Number of Rotary Rigs Running - by State."

In Table 8.11, we give 82 values of the weekly number of rotary rigs in use in the Southern Louisiana Inland Waterways, from the period starting December 10, 1979. A time series plot of this data is shown in Figure 8.24. An informal inspection of this plot reveals that the neighboring observations tend to behave similarly, suggesting a positive correlation between them. This is in contrast to neighboring values alternating in sign implying a negative correlation between them. There does not appear to be a well discernable underlying trend in this data, nor does the data reveal any systematic fluctuations indicative of an underlying periodicity.

In Tables 8.12 and 8.13, we give values of the estimated autocorrelations and partial autocorrelations of this data for lags $1,2,\ldots,25$. In Figures 8.25 and 8.26, we show plots of the estimated autocorrelation function and partial autocorrelation function, respectively. The estimated autocorrelation function appears to decay exponentially, and the estimated partial autocorrelation function shows a value which is significantly differently from zero at lag 1 only. These plots suggest that the data of Table 8.11 may be reasonably well described by an autoregressive process of order 1.

An estimate of the autoregressive parameter β_1 can be obtained by using the estimate $r(1) = .722$ in (8.28). The estimate $r(1)$ is consistent with our observation that the neighboring values in Figure 8.24 tend to behave similarly.

Based upon the above, a proposed linear stochastic model for the weekly number of rotary rigs in use in Southern Louisiana Inland Waterways

Table 8.11

Values of the Weekly Rotary Rigs in Use in
The Southern Louisiana Inland Waterways for the
period starting December 10, 1979

Week Number	Rigs In Use	Week Number	Rigs In Use	Week Number	Rigs In Use	Week Number	Rigs In Use
1	74	21	75	41	75	61	81
2	79	22	84	42	80	62	79
3	82	23	83	43	78	63	75
4	83	24	76	44	79	64	77
5	84	25	80	45	81	65	77
6	79	26	79	46	82	66	79
7	80	27	74	47	80	67	87
8	78	28	73	48	79	68	88
9	75	29	70	49	73	69	87
10	72	30	70	50	75	70	90
11	73	31	68	51	76	71	99
12	74	32	68	52	78	72	88
13	77	33	74	53	81	73	84
14	84	34	68	54	79	74	79
15	81	35	73	55	83	75	82
16	77	36	72	56	82	76	84
17	76	37	74	57	81	77	81
18	72	38	74	58	77	78	75
19	77	39	79	59	75	79	83
20	73	40	81	60	74	80	84
						81	84
						82	81

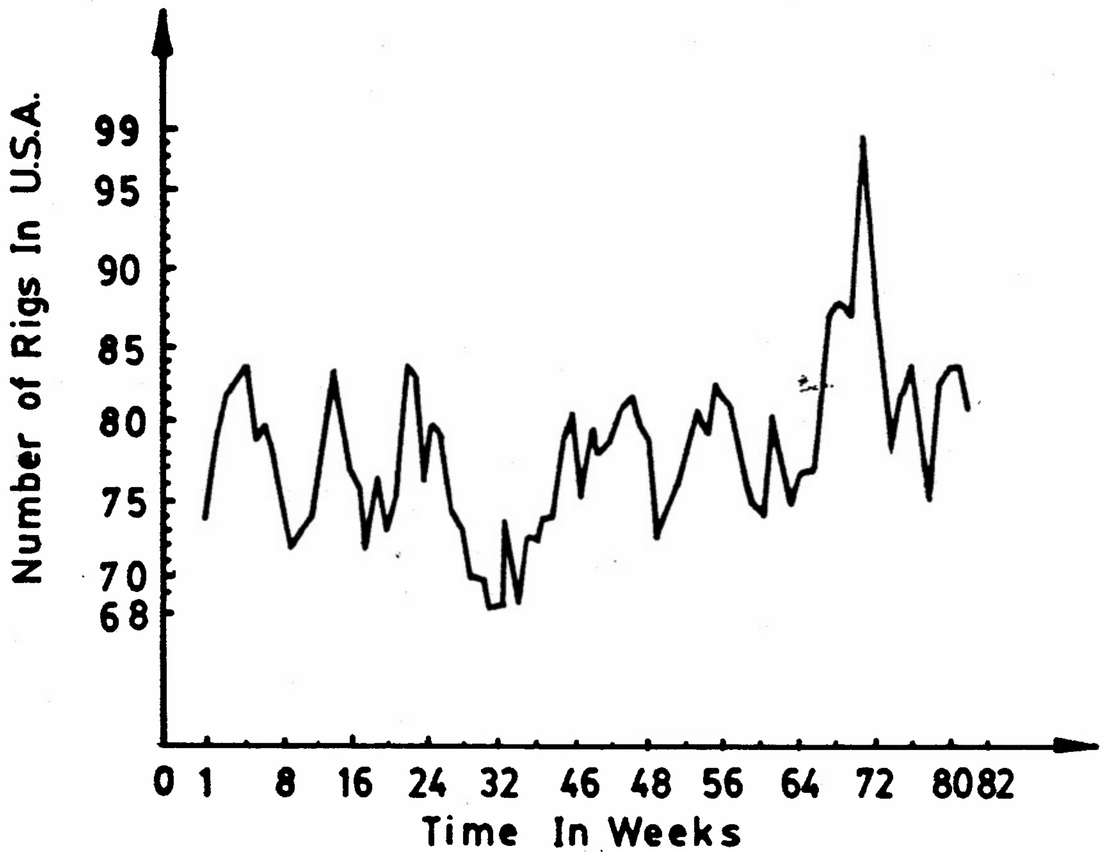

Figure 8.24. A plot of the Weekly Rotary Rigs in Use in The Southern Louisiana Inland Waterways for the period starting December 10, 1979. (See Table 8.11)

Table 8.12

Values of the estimated autocorrelation function $r(h)$,
for $h = 1,\ldots,25$, of the Weekly Rigs in Use data of Table 8.9

Lag h	1	2	3	4	5	6	7	8	9	10	11	12	
Value of r(R)	.72	.52	.4	.28	.18	.11	.07	.12	.18	.16	.11	.05	
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of r(h)	.04	.09	.08	.06	.09	.08	.07	.06	.00	.00	.04	.03	.04

Table 8.13

Values of the estimated partial autocorrelation function $\hat{\pi}(h)$, $h = 1,\ldots,25$, of the Weekly Rigs in Use data of Table 8.9

Lag h	1	2	3	4	5	6	7	8	9	10	11	12	
Value of $\hat{\pi}(h)$	.72	.00	.04	-.04	-.03	-.02	.02	.15	.08	-.06	-.08	-.08	
Lag h	13	14	15	16	17	18	19	20	21	22	23	24	25
Value of $\hat{\pi}(h)$	.05	.17	-.02	-.03	-.03	-.00	-.02	.05	-.04	.04	.02	-.06	.05

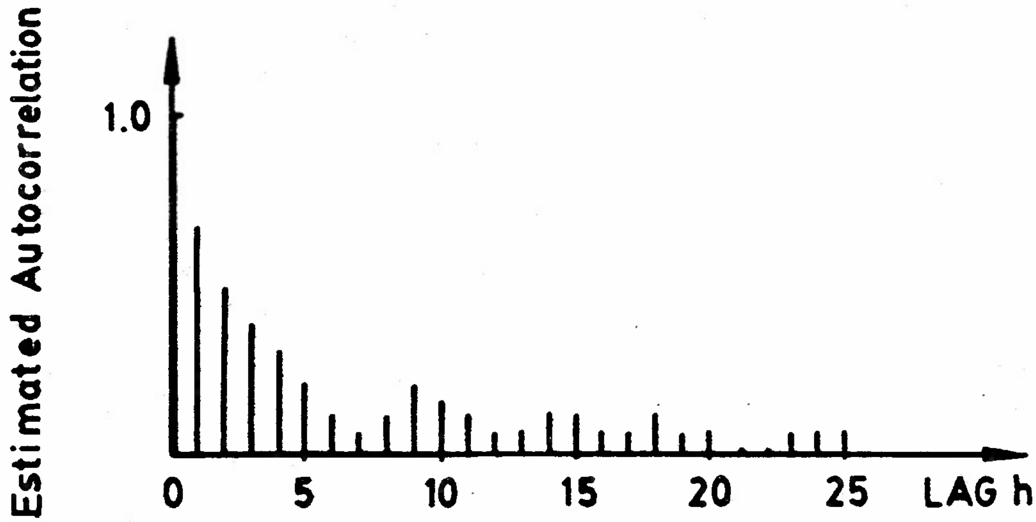

Figure 8.25. A plot of the estimated autocorrelation function of the Weekly Rigs in Use data.

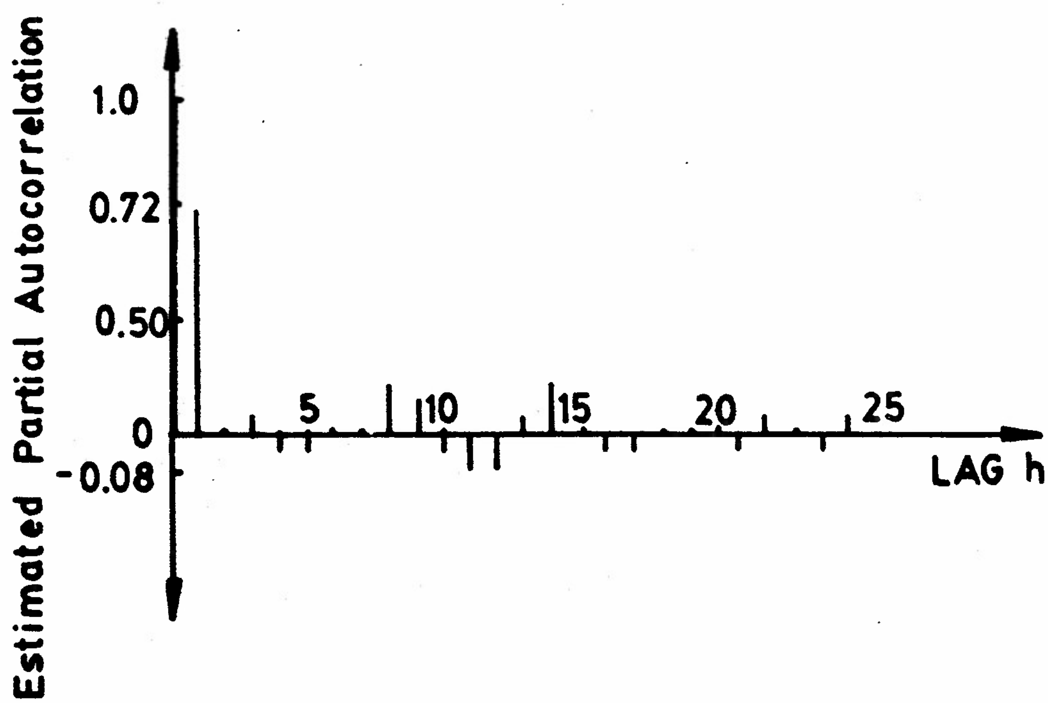

Figure 8.26. A plot of the estimated partial autocorrelation function of the Weekly Rigs in Use data.

y_t is for $t = 2,\ldots,82$

$$(y_t - 78.5) = .722(y_{t-1} - 78.5) + \hat{u}_t \ ,$$

where the $\hat{u}_t$'s , with

$$\hat{u}_t = (y_t - 78.5) - .722(y_{t-1} - 78.5) \ , \qquad t = 2,\ldots,82 \ ,$$

are known as the residuals. The quantity 78.5 represents the mean of the data.

In order to assess how well the proposed mdoel describes the data of Table 8.11, we see if there is any recognizable pattern in the residuals. If the model were adequate, then we would expect that as the series length increases, the $\hat{u}_t$'s would become close to the innovations u_t's . Thus a study of the $\hat{u}_t$'s would indicate the existence and possibly the nature of model inadequacy. In particular, the behavior of the estimated autocorrelation and partial autocorrelation of the $\hat{u}_t$'s would yield valuable evidence about model inadequacy. The absence of a recognizable pattern in a plot of these functions would give us some assurance of model adequacy.

In Tables 8.14 and 8.15, we give values of the estimated autocorrelations and partial autocorrelations of the residuals in question for lags $1,\ldots,25$. In Figures 8.27 and 8.28, we show plots of the data in Tables 8.14 and 8.15. Since these plots do not reveal any recognizable pattern, we conclude the adequacy of the proposed model.

Table 8.14

Values of the estimated autocorrelation function $r(h)$, $h = 1,\ldots,20$, of the residuals from an AR(1) model for the Weekly Rigs in Use data of Table 8.11

Lag h	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	.00	-.02	.06	.03	-.03	-.04	-.11	.01	.12	.07	.02	-.04

Lag h	13	14	15	16	17	18	19	20
Value of r(h)	-.08	.12	.03	-.01	-.02	.07	.00	.07

Table 8.15

Values of the estimated partial autocorrelation function $\hat{\pi}(h)$, $h = 1,\ldots,20$, of the residuals from an AR(1) model for the Weekly Rigs in Use data of Table 8.11

Lag h	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	.00	-.02	.06	.03	-.03	-.04	-.12	.01	.12	.10	.03	-.06
Lag h	13	14	15	16	17	18	19	20				
Value of r(h)	-.12	.10	.06	.05	-.01	.02	-.05	.05				

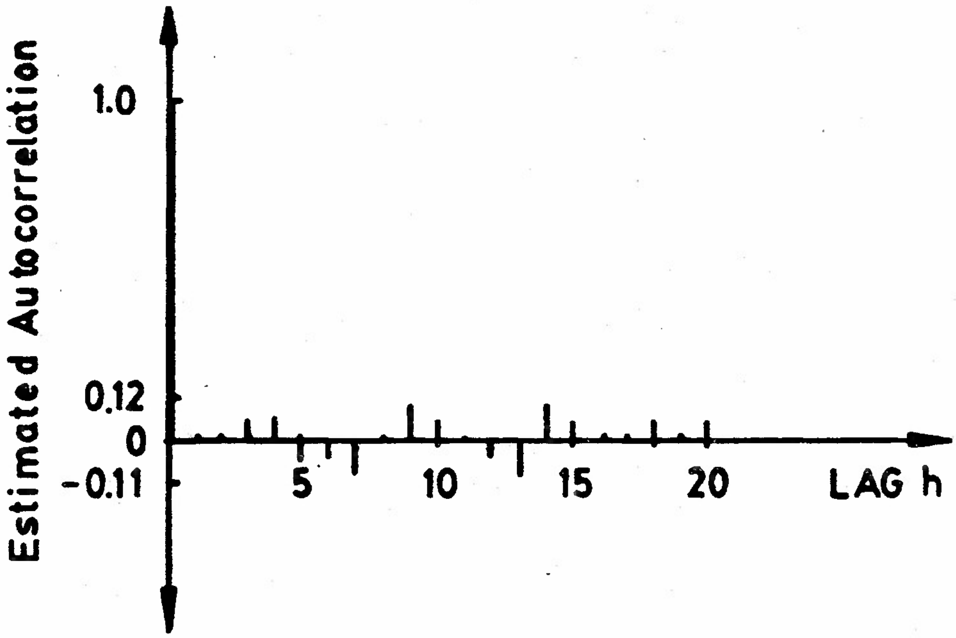

Figure 8.27. A plot of the estimated autocorrelation function of the residuals from an AR(1) model for the Weekly Rigs in Use data.

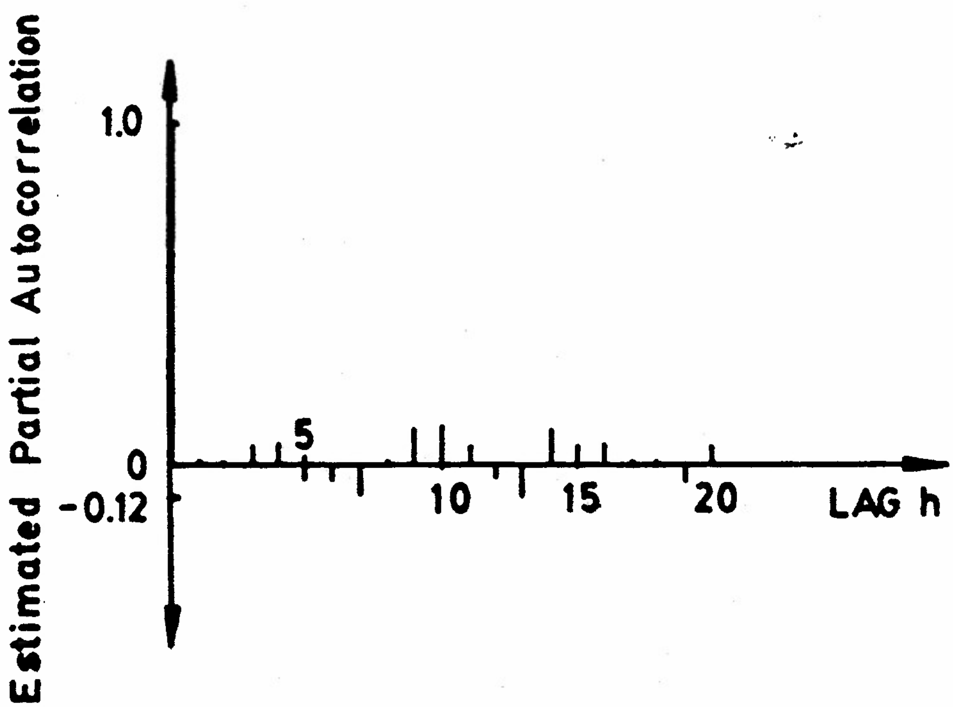

Figure 8.28. A plot of the estimated partial autocorrelation function of the residuals from an AR(1) model for the Weekly Rigs in Use data.

8.11.2 The Landsat 2 Satellite Data*

The Landsat 2 satellite is an earth orbiting satellite which measures the amount of reflected energy in 4 bands of the electromagnetic spectrum. The satellite travels from north to south over the day side of the earth. As the satellite travels, an oscillating mirror sweeps out a 150 kilometer long scan line in a west to east direction. The mirror reflects the energy from the ground onto an array of detectors on board the satellite. The reflected energy is converted to an electrical impulse. Several such impulses are integrated over a short period of time and then transmitted to the earth via ground stations as discrete signals. The discrete signals represent light intensities, 0 denoting black, and large values such as 130 denoting bright.

In Table 8.16 we give 496 values of the light intensities observed by such a satellite over a sand dune field in the Sahara Desert. The measurements are indexed by the distance traveled by the satellite (instead of time) and are recorded at every 80 meter distance. Note that the entries in Table 8.16 are taken from a computer output with an exponent notation; thus the first observation .74000D + 02 denotes 74.000 . The light intensities in this table range from 0 to 127 . In Figure 8.29 we show a plot of the first

---

*This data and its description was given to us by Dr. Mark Labovitz of the National Aeronautics and Space Administration, at the Goddard Space Flight Center, Greenbelt, Maryland 20771.

Table 8.16

Values of the Light Intensities Observed by a Landsat 2 Satellite over the Sahara Desert

1-	8	0.74000D+02	0.71000D+02	0.74000D+02	0.76000D+02	0.68000D+02	0.61000D+02	0.58000D+02	0.53000D+02
9-	16	0.66000D+02	0.68000D+02	0.70000D+02	0.86000D+02	0.94000D+02	0.96000D+02	0.96000D+02	0.92000D+02
17-	24	0.86000D+02	0.79000D+02	0.75000D+02	0.71000D+02	0.75000D+02	0.80000D+02	0.82000D+02	0.87000D+02
25-	32	0.83000D+02	0.78000D+02	0.86000D+02	0.93000D+02	0.93000D+02	0.95000D+02	0.86000D+02	0.92000D+02
33-	40	0.94000D+02	0.10100D+03	0.11000D+03	0.11100D+03	0.10200D+03	0.95000D+02	0.96000D+02	0.95000D+02
41-	48	0.95000D+02	0.10200D+03	0.10100D+03	0.85000D+02	0.78000D+02	0.61000D+02	0.54000D+02	0.54000D+02
49-	56	0.53000D+02	0.70000D+02	0.80000D+02	0.78000D+02	0.75000D+02	0.70000D+02	0.71000D+02	0.71000D+02
57-	64	0.88000D+02	0.95000D+02	0.99000D+02	0.97000D+02	0.93000D+02	0.84000D+02	0.83000D+02	0.79000D+02
65-	72	0.78000D+02	0.77000D+02	0.78000D+02	0.75000D+02	0.77000D+02	0.87000D+02	0.94000D+02	0.10100D+03
73-	80	0.10500D+03	0.10300D+03	0.10000D+03	0.10000D+03	0.93000D+02	0.82000D+02	0.77000D+02	0.79000D+02
81-	88	0.87000D+02	0.86000D+02	0.90000D+02	0.10100D+03	0.10200D+03	0.10000D+03	0.10100D+03	0.10300D+03
89-	96	0.10500D+03	0.10100D+03	0.90000D+02	0.79000D+02	0.70000D+02	0.67000D+02	0.69000D+02	0.73000D+02
97-	104	0.76000D+02	0.74000D+02	0.75000D+02	0.83000D+02	0.87000D+02	0.87000D+02	0.89000D+02	0.82000D+02
105-	112	0.77000D+02	0.84000D+02	0.76000D+02	0.72000D+02	0.72000D+02	0.79000D+02	0.83000D+02	0.75000D+02
113-	120	0.72000D+02	0.68000D+02	0.73000D+02	0.79000D+02	0.89000D+02	0.86000D+02	0.86000D+02	0.92000D+02
121-	128	0.92000D+02	0.90000D+02	0.89000D+02	0.86000D+02	0.77000D+02	0.75000D+02	0.90000D+02	0.71000D+02
129-	136	0.90000D+02	[illegible]	0.93000D+02	0.86000D+02	0.82000D+02	0.77000D+02	0.79000D+02	0.86000D+02
137-	144	0.87000D+02	0.91000D+02	0.91000D+02	0.88000D+02	0.81000D+02	0.84000D+02	0.94000D+02	0.10100D+03
145-	152	0.76000D+02	0.95000D+02	0.76000D+02	0.93000D+02	0.83000D+02	0.76000D+02	0.71000D+02	0.74000D+02
153-	160	0.80000D+02	0.73000D+02	0.70000D+02	0.77000D+02	0.80000D+02	0.84000D+02	0.79000D+02	0.78000D+02
161-	168	0.92000D+02	0.89000D+02	0.84000D+02	0.84000D+02	0.89000D+02	0.95000D+02	0.88000D+02	0.87000D+02
169-	176	0.90000D+02	0.85000D+02	0.86000D+02	0.83000D+02	0.79000D+02	0.77000D+02	0.79000D+02	0.81000D+02
177-	184	0.79000D+02	0.82000D+02	0.83000D+02	0.79000D+02	0.73000D+02	0.69000D+02	0.67000D+02	0.66000D+02
185-	192	0.60000D+02	0.64000D+02	0.72000D+02	0.74000D+02	0.74000D+02	0.69000D+02	0.67000D+02	0.62000D+02
193-	200	0.64000D+02	0.67000D+02	0.73000D+02	0.79000D+02	0.79000D+02	0.85000D+02	0.78000D+02	0.80000D+02
201-	208	0.87000D+02	0.87000D+02	0.90000D+02	0.96000D+02	0.96000D+02	0.10100D+03	0.74000D+02	0.75000D+02
209-	216	0.74000D+02	0.69000D+02	0.78000D+02	0.84000D+02	0.94000D+02	0.96000D+02	0.90000D+02	0.74000D+02
217-	224	0.67000D+02	0.58000D+02	0.69000D+02	0.72000D+02	0.83000D+02	0.84000D+02	0.91000D+02	0.10200D+03
225-	232	0.97000D+02	0.87000D+02	0.89000D+02	0.88000D+02	0.90000D+02	0.82000D+02	0.70000D+02	0.66000D+02
233-	240	0.62000D+02	0.63000D+02	0.59000D+02	0.65000D+02	0.67000D+02	0.67000D+02	0.71000D+02	0.70000D+02
241-	248	0.62000D+02	0.59000D+02	0.56000D+02	0.62000D+02	0.63000D+02	0.71000D+02	0.74000D+02	0.81000D+02
249-	256	0.80000D+02	0.87000D+02	0.91000D+02	0.10000D+03	0.10300D+03	0.98000D+02	0.87000D+02	0.84000D+02
257-	264	0.83000D+02	0.77000D+02	0.71000D+02	0.61000D+02	0.74000D+02	0.87000D+02	0.95000D+02	0.92000D+02
265-	272	0.84000D+02	0.83000D+02	0.82000D+02	0.83000D+02	0.85000D+02	0.84000D+02	0.83000D+02	0.72000D+02
273-	280	0.61000D+02	0.67000D+02	0.79000D+02	0.87000D+02	0.88000D+02	0.83000D+02	0.86000D+02	0.81000D+02
281-	288	0.87000D+02	0.89000D+02	0.88000D+02	0.86000D+02	0.83000D+02	0.84000D+02	0.95000D+02	0.96000D+02
289-	296	0.97000D+02	[illegible]	0.92000D+02	0.91000D+02	0.89000D+02	0.93000D+02	0.91000D+02	0.95000D+02
297-	304	0.99000D+02	0.10100D+03	0.99000D+02	0.10300D+03	0.93000D+02	0.88000D+02	0.91000D+02	0.83000D+02
305-	312	0.80000D+02	0.88000D+02	0.93000D+02	0.97000D+02	0.89000D+02	[illegible]	0.87000D+02	0.83000D+02
313-	320	0.73000D+02	0.10600D+03	0.10800D+03	0.11000D+03	0.10800D+03	0.90000D+02	0.76000D+02	0.70000D+02
321-	328	0.76000D+02	0.80000D+02	0.73000D+02	0.69000D+02	0.62000D+02	0.62000D+02	0.69000D+02	0.70000D+02
329-	336	0.63000D+02	0.70000D+02	0.71000D+02	0.66000D+02	0.65000D+02	0.69000D+02	0.70000D+02	0.70000D+02
337-	344	0.75000D+02	0.85000D+02	0.81000D+02	0.85000D+02	0.84000D+02	0.85000D+02	0.88000D+02	0.86000D+02
345-	352	0.78000D+02	0.74000D+02	0.65000D+02	0.66000D+02	0.78000D+02	0.84000D+02	0.84000D+02	0.83000D+02
353-	360	0.69000D+02	0.68000D+02	0.69000D+02	0.62000D+02	0.50000D+02	0.52000D+02	0.54000D+02	0.62000D+02
361-	368	0.66000D+02	0.69000D+02	0.71000D+02	0.74000D+02	0.81000D+02	0.84000D+02	0.88000D+02	0.88000D+02
369-	376	0.91000D+02	0.10200D+03	0.10000D+03	0.96000D+02	0.92000D+02	0.96000D+02	0.94000D+02	0.89000D+02
377-	384	0.83000D+02	0.75000D+02	0.73000D+02	0.75000D+02	0.80000D+02	0.85000D+02	0.83000D+02	0.89000D+02
385-	392	0.78000D+02	0.91000D+02	0.84000D+02	0.84000D+02	0.82000D+02	0.77000D+02	0.60000D+02	0.58000D+02
393-	400	0.52000D+02	0.54000D+02	0.61000D+02	0.78000D+02	0.90000D+02	0.99000D+02	0.87000D+02	0.84000D+02
401-	408	0.80000D+02	0.80000D+02	0.83000D+02	0.84000D+02	0.77000D+02	0.85000D+02	0.73000D+02	0.90000D+02
409-	416	0.87000D+02	0.83000D+02	0.89000D+02	0.87000D+02	0.76000D+02	0.94000D+02	0.78000D+02	0.75000D+02
417-	424	0.71000D+02	0.73000D+02	0.91000D+02	0.10200D+03	0.10900D+03	0.10700D+03	0.11100D+03	0.11200D+03
425-	432	0.11100D+03	0.10500D+03	0.95000D+02	0.10200D+03	0.99000D+02	0.89000D+02	0.83000D+02	0.81000D+02
433-	440	0.88000D+02	0.90000D+02	0.96000D+02	0.93000D+02	0.97000D+02	0.99000D+02	0.97000D+02	0.93000D+02
441-	448	0.90000D+02	0.84000D+02	0.82000D+02	0.77000D+02	0.81000D+02	0.88000D+02	0.90000D+02	0.89000D+02
449-	456	0.95000D+02	0.84000D+02	0.79000D+02	0.80000D+02	0.86000D+02	0.90000D+02	0.87000D+02	0.99000D+02
457-	464	0.93000D+02	0.96000D+02	0.92000D+02	0.95000D+02	0.82000D+02	0.74000D+02	0.71000D+02	0.64000D+02
465-	472	0.88000D+02	0.81000D+02	0.91000D+02	0.82000D+02	0.79000D+02	0.75000D+02	0.75000D+02	0.74000D+02
473-	480	0.82000D+02	0.71000D+02	0.88000D+02	0.75000D+02	0.10400D+03	0.10200D+03	0.97000D+02	0.99000D+02
481-	488	0.98000D+02	0.94000D+02	0.95000D+02	0.80000D+02	0.74000D+02	0.75000D+02	0.75000D+02	0.77000D+02
489-	496	0.91000D+02	0.88000D+02	0.84000D+02	0.69000D+02	0.68000D+02	0.76000D+02	0.79000D+02	0.74000D+02

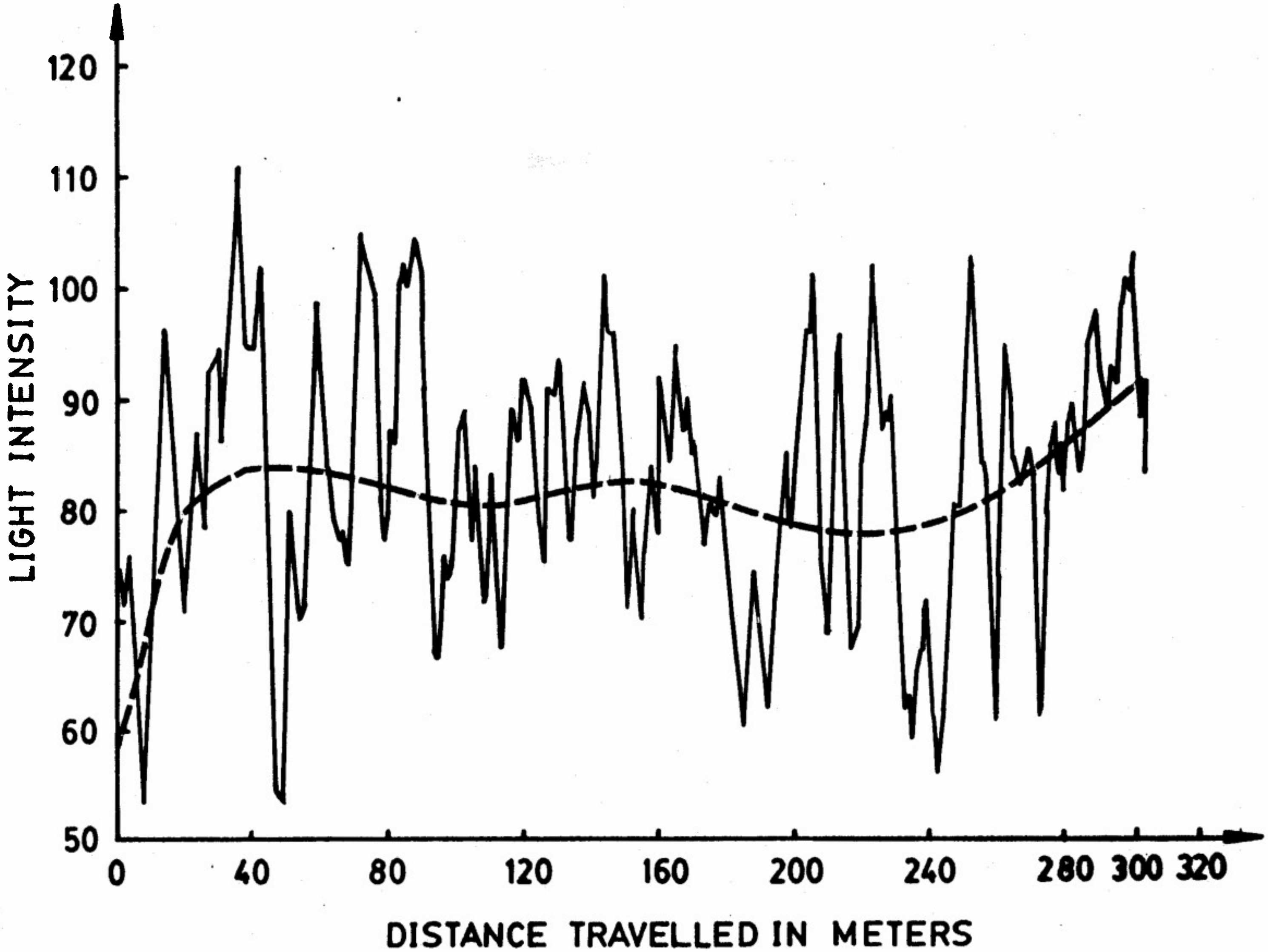

Figure 8.29. A plot of the first 300 values of the light intensity observed by a Landsat 2 Satellite over the Sahara Desert (see Table 8.16).

(The dotted line indicates the nature of a possible underlying trend.)

300 values of the light intensity over distance traveled. The plotting of all the 496 observations would be cumbersome and would tend to conceal the fluctuating behavior of the individual observations. The periodic fluctuations in Figure 8.29 suggest an autoregressive process of order 2 or more - see Section 8.6. The plot also conveys the impression of an underlying trend whose wave like nature is indicated by the dotted line of Figure 8.29. Thus one possibility would be first to fit a cyclical trend of the type described in Section 4.2 to this data, and then to describe the residuals from such a fit by autoregressive processes. However, this possibility was not investigated here, and instead fluctuations about the sample mean of the entire series of 496 observations were considered. The sample mean of the entire set of these data is 82.9.

In Tables 8.17 and 8.18 we give values of the estimated autocorrelations and partial autocorrelations of the deviations of the light intensities from 82.9 for lags $1,2,\ldots,35$. In Figures 8.30 and 8.31 we show plots of the estimated autocorrelation function and partial autocorrelation function, respectively. The estimated autocorrelation function appears to decay exponentially, and perhaps even sinusoidally. The estimated partial autocorrelation function takes values which are significantly different from zero at lags 1 and 2. These plots suggest that the deviations of the light intensities can be described by an autoregressive process of order 2.

Estimates of the two autoregressive parameters β_1 and β_2 can be obtained by using the estimates $r(1) = .87$ and $r(2) = .66$ in

Table 8.17

Values of the estimated autocorrelation function $r(h)$, $h = 1,2,\ldots,35$, of the deviations of the light intensities from their mean for the Landsat 2 satellite data

Lag h	1	2	3	4	5	6	7	8	9	10	11	12
Value of r(h)	.87	.66	.45	.29	.18	.11	.08	.05	.03	.02	.00	.00
Lag h	13	14	15	16	17	18	19	20	21	22	23	24
Value of r(h)	.02	.03	.02	.01	-.02	-.05	-.09	-.09	-.06	-.03	.00	.02
Lag h	25	26	27	28	29	30	31	32	33	34	35	
Value of r(h)	.02	.02	.02	.00	-.02	-.04	-.05	-.06	-.05	-.05	-.05	

Table 8.18

Values of the partial autocorrelation function $\hat{\pi}(h)$, $h = 1,2,\ldots,35$ of the deviations of the light intensities from their mean for the Landsat 2 Satellite data

Lag h	1	2	3	4	5	6	7	8	9	10	11	12
Value of $\hat{\pi}(h)$	.87	-.40	.01	-.02	.05	.02	-.02	-.01	-.01	.00	-.01	.05
Lag h	13	14	15	16	17	18	19	20	21	22	23	24
Value of $\hat{\pi}(h)$	.04	-.05	-.03	-.03	-.01	-.05	-.03	.11	.00	-.01	.02	-.02
Lag h	25	26	27	28	29	30	31	32	33	34	35	
Value of $\hat{\pi}(h)$	.02	-.03	.00	-.07	.03	-.01	-.03	.03	.01	-.04	.01	

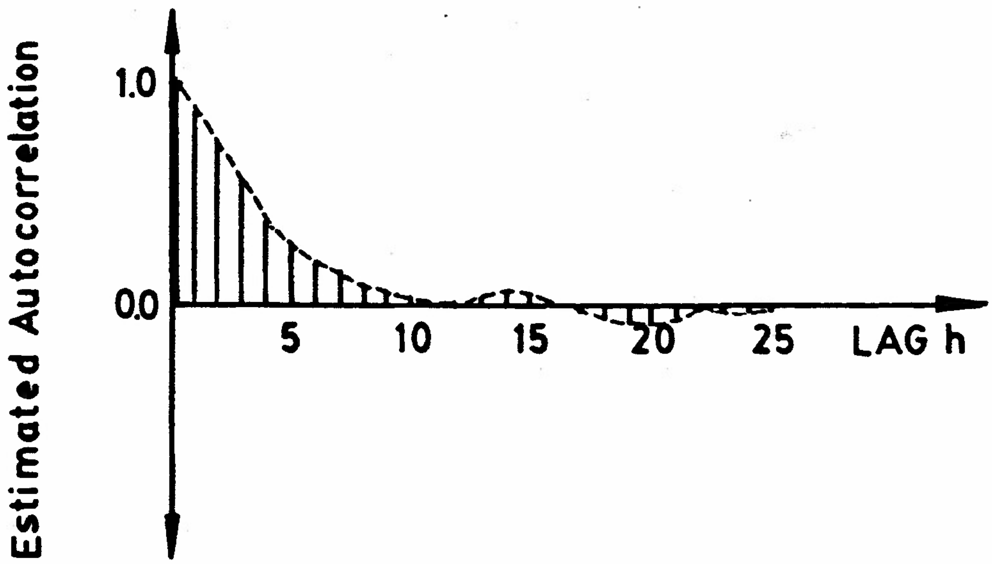

Figure 8.30. A plot of the estimated autocorrelation function of the deviations of the light intensities from their mean for the Landsat 2 Satellite data.

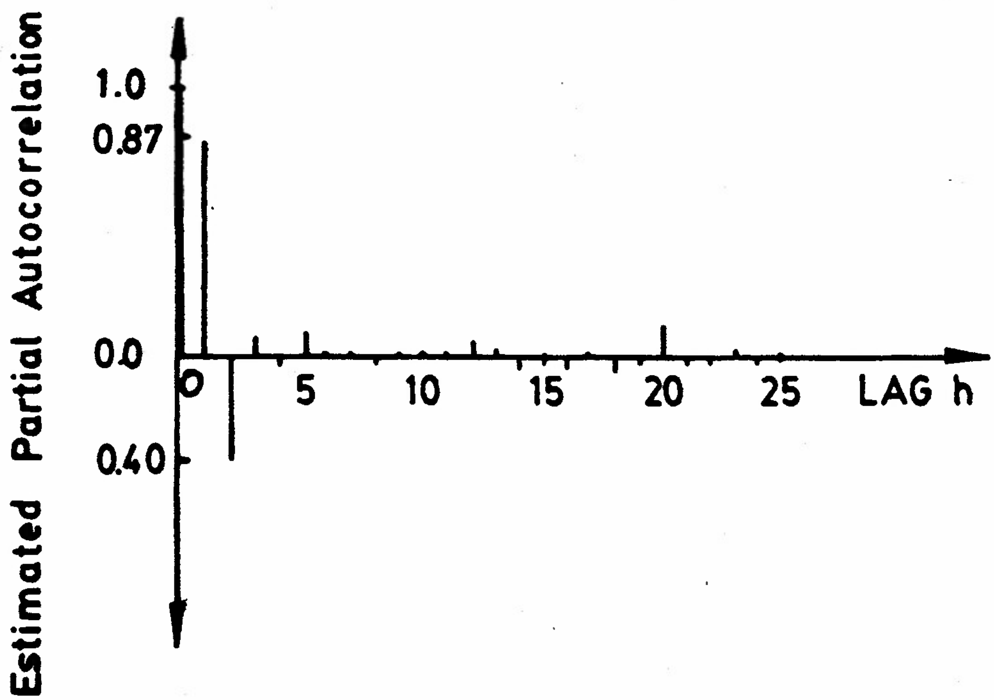

Figure 8.31. A plot of the estimated partial autocorrelation function of the deviations of the light intensities from their mean for the Landsat 2 Satellite data.

Equation (8.28); these turn out to be $\hat{\beta}_1 = -1.28$ and $\hat{\beta}_2 = .46$. The estimate of β_2 is approximately the estimated value of the partial autocorrelation at lag 2, which is $-.40$.

Based upon the above, a proposed linear stochastic model for the light intensity y_t, is, for $t = 3,\ldots,496$, $(y_t - 82.9) = 1.28(y_{t-1}-82.9) - .46(y_{t-2} - 82.9) + \hat{u}_t$, where the $\hat{u}_t$'s, with

$$\hat{u}_t = (y_t - 82.9) - 1.28(y_{t-1} - 82.9) + .46(y_{t-2} - 82.9)\ , \quad t = 3,\ldots,496\ ,$$

are known as the <u>residuals</u>.

In order to judge how well the proposed model describes the data of Table 8.16, we see if there is any recognizable pattern in the residuals. If the model were adequate, then we would expect that as the series length increases the $\hat{u}_t$'s would become close to the innovations u_t's. Thus a study of the $\hat{u}_t$'s would indicate the existence and possibly the nature of model inadequacy. In particular, the behavior of the estimated autocorrelation and partial autocorrelation functions of the $\hat{u}_t$'s would yield valuable evidence about model inadequacy. The absence of a recognizable pattern in a plot of these functions would give us some assurance of model adequacy.

In Tables 8.19 and 8.20 we give values of the estimated autocorrelations and partial autocorrelations of the residuals in question for lags $1,2,\ldots,35$. In Figures 8.32 and 8.33, we show plots of the data in Tables 8.14 and 8.15, respectively. Since these plots do not show any recognizable pattern, we conclude that the proposed model is adequate.

Table 8.19

Values of the estimated autocorrelation function $r(h)$, $h = 1,2,\ldots,35$, of the residuals from an AR(2) model for the Landsat 2 Satellite data

Lag h	1	2	3	4	5	6	7	8	9	10	11	12
Values of r(h)	.03	-.05	.04	-.04	.00	.02	.04	.02	.00	.04	-.02	-.06
Lag h	13	14	15	16	17	18	19	20	21	22	23	24
Values of r(h)	.04	.02	.04	-.02	.01	.05	-.02	-.05	.01	-.02	.03	0.00
Lag h	25	26	27	28	29	30	31	32	33	34	35	
Values of r(h)	.02	-.03	.07	-.01	-.03	.01	-.02	-.05	.02	.00	.01	

Table 8.20

Values of the estimated partial autocorrelation function $\hat{\pi}(h)$, $h = 1,2,\ldots,35$, of the residuals from an AR(2) model for the Landsat 2 Satellite data

Lag h	1	2	3	4	5	6	7	8	9	10	11	12
Values of r(h)	.03	-.06	.04	-.04	.01	.02	.04	.02	0.00	.04	-.02	-.05
Lag h	13	14	15	16	17	18	19	20	21	22	23	24
Values of r(h)	.03	.02	.04	-.03	.01	.05	-.02	-.04	.00	-.01	.02	-.02
Lag h	25	26	27	28	29	30	31	32	33	34	35	
Values of r(h)	.03	-.03	.09	-.04	.00	.01	-.03	-.05	.02	.01	.01	

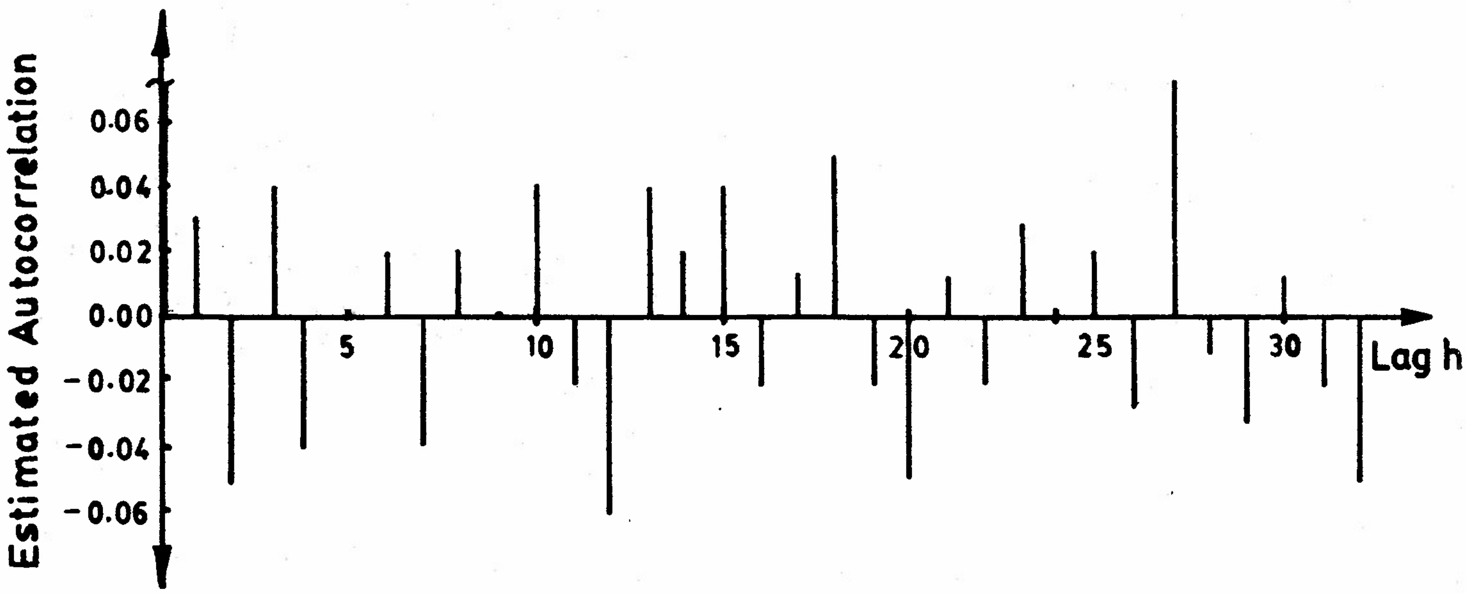

Figure 8.32. A plot of the estimated autocorrelation function of the residuals from an AR(2) model for the Landsat 2 Satellite data.

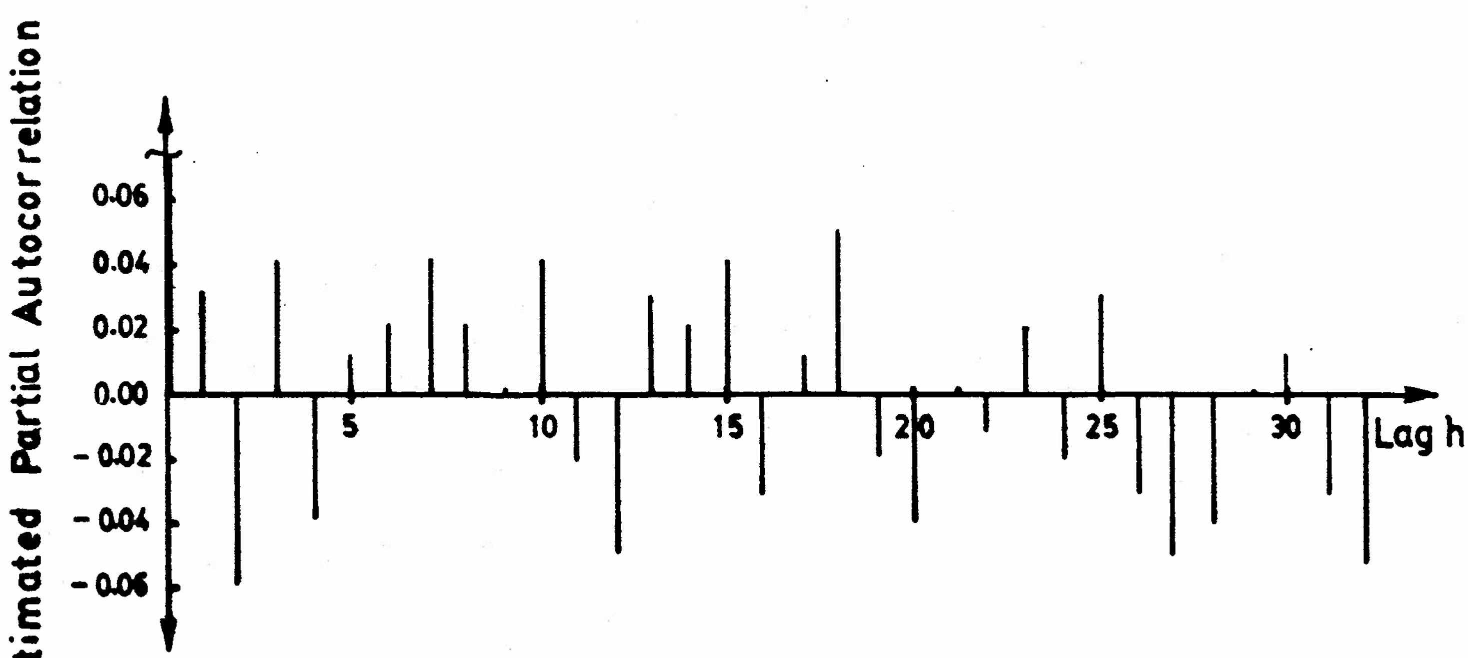

Figure 8.33. A plot of the estimated partial autocorrelation function of the residuals from an AR(2) model for the Landsat 2 Satellite data.

References

Anderson, T. W. (1971). The Statistical Analysis of Time Series. New York: John Wiley & Sons, Inc.

Anderson, T. W. (1984). An Introduction to Multivariate Statistical Analysis, Second Edition. New York: John Wiley & Sons, Inc.

Box, G.E.P., and Jenkins, G.M. (1976). Time Series Analysis: forecasting and control. San Francisco: Holden-Day.

Walker, G. (1931). On periodicity in series of related terms, Proc. Roy. Soc. London Ser. A, 131, 518-532.

Yule, G. U. (1927). On a method for investigating periodicities in disturbed series with special reference to Wolfer's sunspot numbers, Philos. Trans. Roy. Soc. London Ser. A, 226, 237-298.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 12	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Methods and Applications of Time Series Analysis Part II: Linear Stochastic Models		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) T. W. Anderson and N. D. Singpurwalla		8. CONTRACT OR GRANT NUMBER(s) DAAG 29-82-K-0156
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Statistics - Sequoia Hall, Stanford University, Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P-19065-M
11. CONTROLLING OFFICE NAME AND ADDRESS U.S. Army Research Office, Post Office Box 12211, Research Triangle Park, NC 27709		12. REPORT DATE October 1984
		13. NUMBER OF PAGES 120
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the authors and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Time series analysis, autoregressive processes, moving average representation, autocorrelation function, partial autocorrelation function, forecasting

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

This is the second in a series of technical reports developing the most modern procedures of time series analysis and forecasting for use in engineering, the physical sciences, and the social sciences. The exposition of methodology is based on a succinct presentation of the theoretical background and is illustrated with appropriate examples from engineering, maintenance and reliability, economics, and other physical and social sciences. This report is concerned with linear statistical models, especially autoregressive processes.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)